Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin

MULTILEVEL GRADED RESPONSE MODELS FOR ANALYZING LONGITUDINAL MULTITRAIT-MULTIMETHOD DATA

Dissertation
zur Erlangung des akademischen Grades
Doktor(in) der Philosophie (Dr. phil.)

vorgelegt von
M. Sc. M. Sc.

Jana Holtmann

Berlin, 2017

Primary Advisor: Prof. Dr. Michael Eid
(Erstgutachter)

Secondary Advisor: Prof. Dr. Fridtjof Nussbeck
(Zweitgutachter)

Date of Defense: 20.10.2017
(Datum der Verteidigung)

Acknowledgements

I would like to thank: *Michael Eid* for his excellent mentoring, good advices, and availability for discussions and questions; *Fridtjof Nussbeck* for willing to invest time and effort to read and review this rather dry work; *Tobias Koch* for his support and patience in answering my questions; *Johannes Bohn*, my fellow PhD student in the project, for the good cooperation and his good work in data collection; *Esther Ulitzsch* and *Louisa Hohmann* for invaluable help in generating latex tables and proofreading parts of this work; *Georg Hosoya* for sharing his enthusiasm for Bayesian data analysis; my colleagues *Claudia Crayen, Martin Schultze, Georg Hosoya, Jana Mahlke, Tanja Kutscher*, and *Johannes Bohn* for brightening up my lunch breaks and helping me through times of low motivation; and of course my family and friends for their invaluable support.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit allein verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Diese Arbeit ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Berlin,

Jana Holtmann

Contributions

The work presented on the LS-Com graded response model has been pre-published in:

Holtmann, J., Koch, T., Bohn, J., & Eid, M. (2017). Bayesian analysis of longitudinal mulitrait-multimethod data with ordinal response variables. *British Journal of Mathematical and Statistical Psychology, 70*, 42 - 80. doi: 10.1111/bmsp.12081

This concerns Sections 1.1 - 1.3, 2.1 - 2.3, and 7.5. Also, the application of the LS-Com graded response model to a smaller subset (with fewer time points and methods) of the well-being data presented here has already been published in the same work. This pre-publication concerns Sections 8.1 - 8.3.

My contributions for the original publication were:

Data analysis: entirely; programming: entirely; compiling of the manuscript: largely; development of methods: largely; reasoning: largely; literature research: largely; discussion of results: entirely; data collection: partially; original idea: partially.

Some of the proofs presented in Section 4.8 have been pre-published for the continuous-indicator LST-Com model in the online supplemental material of:

Koch, T., Schultze, M., Holtmann, J., Geiser, C., & Eid, M. (2017). A multimethod latent state-trait model for structurally different and interchangeable methods. *Psychometrika*, 82, 17 - 47. doi: 10.1007/s11336-016-9541-x

My contribution concerning the proofs in the supplemental material of the original publication was: compiling of the manuscript: largely; reasoning: largely.

Abstract

Investigating and understanding the stability, variability and change of psychological constructs is a major goal in longitudinal psychological assessment. It has been suggested that for a complete understanding of a longitudinal process under investigation, it is crucial to apply multimethod research designs (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Since Campbell and Fiske (1959) it is widely acknowledged that psychological constructs are always assessed using a specific method of observations, and an observation does not only reflect the psychological construct under consideration but does also contain systematic method-specific influences. Multitrait-multimethod (MTMM) designs allow researchers to explicitly model method effects and analyze convergent and discriminant validity of a construct. Despite the growing interest in longitudinal and MTMM data analysis, only few attempts have been made to combine sophisticated longitudinal latent variable models and MTMM data analysis.

To successfully apply longitudinal CFA-MTMM models in practice, it is important to consider specific aspects of the measurement design. First, an increasing number of MTMM measurement designs include a combination of different methods (e.g., different types of raters). Eid et al. (2008) provided a typology of CFA-MTMM models for interchangeable methods, structurally different methods, and a combination of both types of methods. Interchangeable methods are methods that are randomly selected from the same set of methods (e.g., raters). As interchangeable raters are drawn in a multi-stage sampling procedure, the resulting multilevel structure has to be modeled adequately. In contrast, structurally different methods are not selected from the same set of methods and can therefore not be easily replaced by one another (e.g., self-ratings). Until now, only few CFA-MTMM models have been presented allowing researchers to analyze longitudinal MTMM data with structurally different and interchangeable methods (Koch, 2013; Koch, Schultze, Eid, & Geiser, 2014; Koch, Schultze, Holtmann, Geiser, & Eid, 2017).

Second, in longitudinal research, an increasing number of psychological constructs are assessed using short-scales in large-scale panel studies, with an associated increase in the need for models that allow analyses on the item-level. As items are commonly measured on a categorical response scale, measurement models of item response theory (IRT) have to be considered to properly model the response format. Thus far, only few models have been presented allowing researchers to analyze complex MTMM data with ordered response variables (Crayen, Geiser, Scheithauer, & Eid, 2011; Eid, 1996; Jeon & Rijmen, 2014; Nussbeck, Eid, & Lischetzke, 2006), but none of these models can be used for longitudinal MTMM measurement designs combining structurally different and interchangeable methods. The present work fills this gap by introducing several longitudinal multilevel CFA-MTMM models for ordered response variables: a latent state (LS-Com), a latent change (LC-Com), a latent state-trait (LST-Com), and a latent growth curve (LGC-Com) graded response model (GRM). These longitudinal latent variable models belong to the most widely applied CFA approaches to longitudinal data modeling and serve to answer different research questions. The presented models combine the advantages of multilevel MTMM measurement designs and longitudinal CFA models for categorical observed variables.

The complexity of these models with several latent variables and ordinal indicators exceed computational and practical limitations of numerical integration. Presently, only Bayesian estimation methods allow for the estimation of the models proposed in this work.

The statistical performance of the models is investigated via three simulation studies using Bayesian estimation techniques. As the results of the simulation studies show, the LS-Com GRM and LST-Com GRM can be accurately estimated under realistic sample sizes if low degrees of convergent validity are present. These results are encouraging and suggest that even complex multilevel longitudinal CFA-MTMM models can be applied in a wide range of situations using Bayesian methods. However, estimation of the models reaches its limits in cases of high convergent validity and for the LGC-Com GRM with small slope variances. The results of the simulation studies are discussed and practical guidelines for empirical applications are given. An application of the models to multi-rater data on life satisfaction and subjective happiness illustrates the applicability and advantages of the models in applied research as well as the advantages of sampling the model coefficients by Bayesian MCMC methods. Finally, the advantages and limitations of the models are discussed and an outlook on future research topics is provided.

Zusammenfassung

Die Untersuchung und Erklärung der Stabilität, Variabilität und Veränderung psychologischer Konstrukte ist ein wichtiges Ziel längsschnitlicher psychologischer Forschung. Für ein umfassendes Verständnis des zu untersuchenden längsschnittlichen Prozesses wurde dem Einsatz multimethodaler Forschungsdesigns äußerste Wichtigkeit zugesprochen (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Seit Campbell and Fiske (1959) ist die Idee allgemein anerkannt, dass psychologische Konstrukte immer mit einer spezifischen Beobachtungsmethode gemessen werden und somit Beobachtungen neben dem relevanten, zu messenden psychologischen Konstrukt auch systematische methoden-spezifische Einflüsse erfassen. Multitrait-multimethod (MTMM) Designs ermöglichen es, solche Methodeneffekte explizit zu modellieren und die konvergente und diskriminante Validität eines Konstruktes zu analysieren. Trotz des steigenden Interesses an längsschnittlichen sowie an MTMM Datenanalysen wurden nur wenige Versuche unternommen, anspruchsvolle längsschnittliche Modelle für latente Variablen und MTMM Analysen miteinander zu kombinieren.

Für die erfolgreiche Anwendung längsschnittlicher CFA-MTMM Modelle in der Praxis ist es von zentraler Bedeutung, Aspekte des Messdesigns zu berücksichtigen. Zum einen umfasst eine steigende Anzahl von MTMM Messdesigns eine Kombination verschiedener Methoden (z.B. verschiedene Rater-Typen). Eid et al. (2008) erstellten eine Typologie von CFA-MTMM Modellen für austauschbare, strukturell verschiedene, sowie die Kombination beider Typen von Methoden. Austauschbare Methoden sind Methoden (z.B. Rater), welche zufällig aus der gleichen Menge von Methoden gezogen werden. Da austauschbare Rater in einem mehrstufigen Prozess der Stichprobenziehung gewonnen werden, muss die entstehende Multilevel-Struktur der Daten adäquat modelliert werden. Strukturell verschiedene Methoden hingegen werden nicht aus einer Menge gleicher Methoden gezogen und könenn daher nicht einfach durch einander ersetzt werden (z.B. Selbstberichte). Bisher gibt es nur wenige CFA-MTMM Modelle, welche es ermöglichen längsschnittliche MTMM Daten mit einer Kombination von strukturell verschiedenen und austauschbaren Methoden zu analysieren (Koch, 2013; Koch, Schultze, Eid, & Geiser, 2014; Koch, Schultze, Holtmann, Geiser, & Eid, 2017).

Zweitens wird eine steigende Zahl von psychologischen Konstrukten in Panel-Studien anhand von Kurzskalen erhoben, womit der Bedarf an Modellen, welche Analysen auf der Item-Ebene erlauben, wächst. Da Items häufig mit kategorialen Antwortformaten erfasst werden ist es entscheidend dieses kategoriale Antwortformat durch die Verwendung von Messmodellen der Item-Response-Theorie angemessen zu berücksichtigen. Bisher wurden nur wenige Modelle für die Analyse komplexer MTMM Daten mit geordnet kategorialen Antwortvariablen eingeführt (Crayen, Geiser, Scheithauer, & Eid, 2011; Eid, 1996; Jeon & Rijmen, 2014; Nussbeck, Eid, & Lischetzke, 2006). Keines dieser Modelle kann je-

doch für länggschnittliche MTMM Messmodelle mit einer Kombination von austauschbaren und strukturell verschiedenen Methoden angewendet werden. Die vorgelegte Arbeit schließt diese Lücke und präsentiert mehrere längsschnittliche multilevel CFA-MTMM Modelle für geordnet kategoriale Antwortvariablen: ein Latent State (LS-Com), ein Latent Change (LC-Com), ein Latent State-Trait (LST-Com), und ein Latent Growth Curve (LGC-Com) Graded Response Modell (GRM). Diese längsschnittlichen latenten-Variablen-Modelle gehören zu den weit verbreitesten CFA Ansätzen längsschnittlicher Datenanalyse und können zur Beantwortung verschiedener Forschungsfragen herangezogen werden. Die eingeführten Modelle kombinieren die Vorteile von multilevel MTMM Messdesigns und längsschnittlichen CFA Modellen für kategoriale beobachtete Variablen.
Die Komplexität der eingeführten Modelle mit mehreren latenten Variablen und ordinalen Indikatoren überschreitet die Grenzen der Anwendbarkeit und Rechenkapazitäten von Verfahren der numerischen Integration. Folglich können die präsentierten Modelle bisher nur mit Bayesianischen Methoden geschätzt werden.
Die statistische Performanz der Modelle wurde in drei Simulationsstudien mithilfe Bayesianischer Schätzverfahren untersucht. Die Ergebnisse der Simulationsstudien zeigen, dass das LS-Com GRM und das LST-Com GRM unter realistischen Stichprobengrößen akkurat geschätzt werden können, wenn ein moderates Level konvergenter Validität vorliegt. Die Ergebnisse zeigen, dass solch komplexe längsschnittliche multilevel CFA-MTMM Modelle in einer breiten Zahl von Situationen mithilfe Bayesianischer Schätzmethoden angewendet werden können. Die Schätzbarkeit der Modelle stößt jedoch an ihre Grenzen wenn niedrige Level konvergenter Validität vorliegen oder wenn die Slope Varianzen im LGC-Com GRM gering sind.
Die Ergebnisse der Simulationsstudien werden diskutiert und praktische Anwendungsrichtlinien werden vorgestellt. Eine Anwendung der Modelle auf Multi-Rater Daten von subjektiver Happiness und Lebenszufriedenheit illustriert die Anwendbarkeit und die Vorteile der Modelle in angewandter Forschung sowie die Vorteile der Modellschätzung mittels Bayesianischer MCMC Verfahren. Vorteile und Grenzen der Modelle werden diskutiert und ein Ausblick auf zukünftige Forschungsfragen wird gegeben.

Contents

List of Figures

List of Tables

Acronyms and Indices

Acronyms

AIC	Akaike information criterion
BIC	Bayesian information criterion
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Credibility interval
CM	Common method
GRM	Graded response model
ICC	Intraclass correlation coefficient
IRT	Item response theory
LC	Latent change
LGC	Latent growth curve
LS	Latent state
LST	Latent state-trait
MCMC	Markov chain Monte Carlo
MI	Measurement invariance
ML	Maximum likelihood
MSE	Mean squared error
MTMM	Multitrait-multimethod
nL1	number of level-1 units
nL2	number of level-2 units
Peb	Parameter estimation bias
PPP	Posterior predictive p-value

PSR	Potential scale reduction
RMSEA	Root mean square error of approximation
Seb	Standard error bias
SEM	Structural equation modeling / model
SHS	Subjective happiness scale
SRMR	Standardized root mean square residual
SWB	Subjective well-being
SWLS	Satisfaction with life scale
TMU	Trait method unit
UM	Unique method
WLS	Weighted least squares
WLSMV	Weighted least squares means and variance adjusted

Indices

r	Rater
t	Target
i	Indicator
j	Construct
k	Method
l	Measurement Occasion
s	Response category
q_{ij}	Number of response categories of indicator i of construct j

Chapter 1

Introduction

1.1 Longitudinal multitrait-multimethod modeling

The analysis of the stability and change of interindividual differences is one of the most fundamental goals in psychology[1]. Today, a plethora of statistical models are available for the examination of intra- and interindividual change using manifest or latent variable approaches (Heck, Thomas, & Tabata, 2013; Little, Schnabel, & Baumert, 2000; Rabe-Hesketh & Skrondal, 2004; Singer & Willett, 2003; Steele, 2008).

Since Campbell and Fiske first introduced multitrait-multimethod (MTMM) analysis into the social sciences in 1959, it has been repeatedly suggested that researchers should use multimethod designs (Eid & Diener, 2006). Multimethod designs enable researchers to scrutinize important properties of their measures (e.g., convergent and discriminant validity) and investigate method effects. This investigation is essential, as, how Fiske and Campbell (1992) have stated, "Method and trait or content are highly interactive and interdependent". To evaluate the convergent and discriminant validity, at least two methods have to be considered (Campbell & Fiske, 1959; Eid & Diener, 2006). According to Campbell and Fiske (1959, p. 394), convergent validity can be investigated through associations between two distinct methods (e.g., self-reports and parent reports) assessing the same construct or attribute (e.g., happiness). Discriminant validity refers to the associations between two methods assessing different constructs or attributes.

Presently, latent variable approaches such as confirmatory factor analysis (CFA) are commonly used to analyze MTMM data (Eid, 2000; Eid & Diener, 2004; Eid et al., 2003, 2008). Latent variable models allow researchers to model MTMM structures in more sophisticated and versatile ways. In this framework, method effects are not regarded as errors or nuisance parameters but as an integral part of a psychological measure that is worth to be investigated. Numerous CFA-MTMM models have been proposed so far, including traditional and, more recently, design-oriented modeling approaches (Eid, Geiser, & Koch, 2016). Examples of traditional MTMM models are the correlated trait-correlated uniqueness model (CT-CU; Kenny, 1976), the correlated trait-uncorrelated method model (CT-UM; Marsh & Grayson, 1995), and the correlated trait-correlated method model (CT-CM; Marsh & Grayson, 1995). Current developments in MTMM research focus on specific aspects of the measurement design in order to formulate adequate statistical models. Examples of such design-oriented

[1] Note that parts of this introduction have been pre-published in Holtmann, Koch, Bohn, and Eid (2017).

CFA-MTMM models are the CTC(M-1) model (Eid, 2000), the latent difference model (Pohl, Steyer, & Kraus, 2008), or the latent means model (Pohl & Steyer, 2010).
Despite the growing interest in longitudinal and MTMM data analysis, only few attempts have been made to combine sophisticated longitudinal latent variable models and MTMM data analysis (Courvoisier, Nussbeck, Eid, Geiser, & Cole, 2008; Geiser, Eid, Nussbeck, Courvoisier, & Cole, 2010; Grimm, Pianta, & Konold, 2009; Kenny & Zautra, 2001; Koch, 2013; Koch et al., 2014). However, longitudinal CFA-MTMM models bear many advantages (Geiser et al., 2010; Koch et al., 2014). For example, they allow researchers to:

1. explicitly model measurement error and investigate inter- and intraindividual change apart from measurement error influences,

2. evaluate the convergent and discriminant validity of different measures within and across occasions of measurement and thereby evaluate construct validity over time,

3. specify method effects as latent variables and study the change and stability with regard to construct and method effects,

4. test important assumptions of longitudinal data analysis such as measurement invariance, and

5. relate external (explanatory or criterion) variables to the latent variables in the model.

To successfully apply longitudinal CFA-MTMM models in practice, it is important to consider specific aspects of the measurement design. First, an increasing number of MTMM measurement designs include a combination of different methods (e.g., different types of raters). Second, in many empirical applications researchers use single items as indicators. Because items are typically measured on categorical response scales, measurement models of item response theory (IRT) have to be considered.
A design-oriented approach to MTMM modeling requires researchers to consider specific aspects of their measurement design in order to formulate an appropriate model. For that purpose, Eid et al. (2008) provided a typology of CFA-MTMM models for interchangeable methods, structurally different methods, and a combination of both types of methods. Interchangeable methods are methods that are randomly selected from the same set of methods. Consider, for example, multiple peer ratings of students' empathy or well-being. Because multiple peer ratings stem from the same pool of methods (e.g., raters), they will have a similar access to the target's behavior (Eid et al., 2008). As interchangeable methods are drawn in a multi-stage sampling procedure, the resulting multilevel structure has to be modeled adequately.
In contrast, structurally different methods are not selected from the same set of methods and can therefore not be easily replaced by one another. For example, self-ratings, parent ratings, and physiological measures can be considered structurally different methods, as they stem from different sets of methods and often reflect different perspectives on the target's behavior. Thus, the distinction between interchangeable and structurally different methods can be understood in analogy to the distinction between random and fixed effects.
Many of today's MTMM measurement designs incorporate a combination of structurally different and interchangeable methods. For example, a combination of structurally different and interchangeable methods is frequently encountered in educational studies (Eid et al.,

2008; Koch et al., 2016; Pham et al., 2012), organizational studies (Mahlke et al., 2016), and studies from social and personality psychology (Carretero-Dios, Eid, & Ruch, 2011).
Until now, only few CFA-MTMM models have been presented allowing researchers to analyze longitudinal MTMM data with structurally different and interchangeable methods (Koch, 2013; Koch, Schultze, Eid, & Geiser, 2014; Koch, Schultze, Holtmann, Geiser, & Eid, 2017). The model by Koch et al. (2014) can be seen as a longitudinal variant of the original multilevel CFA-MTMM model for structurally different and interchangeable methods proposed by Eid et al. (2008). Despite their advantages, these models are limited to continuous observed variables and cannot be used for single item analysis. In the present work, the models by Koch (2013) are extended to ordinal response variables.

1.2 Analyzing longitudinal MTMM data with ordered response variables

Several reasons call for models that allow to analyze MTMM data with ordered response variables. In recent years, an increasing number of psychological instruments are being used in large-scale social surveys (Rammstedt & Beierlein, 2015). Large-scale panel studies are costly and face economic constraints, requiring time-efficient and short assessments of any construct of interest. Additionally, survey length has been found to be negatively associated with response rates (Edwards, Roberts, Sandercock, & Frost, 2004). Hence, the use of short-scales has become highly relevant in longitudinal survey studies, such as the German Socio-Economic Panel (SOEP; Wagner, Frick, & Schupp, 2007; see, e.g., Boyce, Wood, & Brown, 2010; Headey, Muffels, & Wagner, 2013; Lang, Weiss, Stocker, & von Rosenbladt, 2007; S. M. Schneider & Schupp, 2014, for example studies using psychological short-scales of the SOEP). Using short-scales, however, precludes analyses on a scale level. Instead, researchers are forced to conduct item-level analyses, which is the preferred approach when only few items are available. An appropriate treatment of the categorical response format in item-level analyses is essential, as treating categorical variables as continuous can result in biased parameter estimates as well as incorrect standard errors and test statistics (Beauducel & Herzberg, 2006; Dolan, 1994; Moshagen & Musch, 2014; Rhemtulla, Brosseau-Liard, & Savalei, 2012). Using single items as indicators has further advantages. In longitudinal studies, researchers should ensure that a measure assesses the same construct in the same way across the different measurement occasions. This refers to the concept of measurement (or factorial) invariance (Meredith, 1993; Meredith & Teresi, 2006). To test for factorial invariance, it has been suggested to use single items as indicators (Meredith, 1993). Furthermore, the comparison and selection of items demands a detailed examination of their psychometric properties (e.g., consistencies, stabilities, reliabilities or difficulties), which cannot be investigated if items are aggregated to test parcels.
Thus far, only few models have been presented allowing researchers to analyze MTMM data with ordered response variables (Crayen, Geiser, Scheithauer, & Eid, 2011; Eid, 1996; Jeon & Rijmen, 2014; Nussbeck, Eid, & Lischetzke, 2006). However, none of these models can be used for longitudinal MTMM measurement designs combining structurally different and interchangeable methods. The present work fills this gap by introducing a longitudinal multilevel CFA-MTMM model for ordered response variables.

1.3 The need for and possibilities of analyzing MTMM data with Bayesian estimation techniques

Bayesian analysis facilitates the estimation of highly complex models, such as multilevel structural equation models with categorical response variables (Asparouhov & Muthén, 2010b; Fox, 2005; B. Muthén & Asparouhov, 2012). For these, the use of maximum likelihood estimation would require high-dimensional numerical integration, and the complexity of models with several latent variables and ordinal indicators often exceeds computational and practical limitations of numerical integration. Accordingly, only Bayesian methods allow for the estimation of the models proposed in this work.

Additionally, Bayesian methods bear some advantages for the analysis of MTMM data. First, MTMM studies that include the ratings of several interchangeable raters per target often encounter the problem of small sample sizes. For example, in psychological multirater studies, often two to 10 interchangeable raters are collected for each target-person. Studies have shown that Bayesian methods outperform classical estimation methods (e.g., maximum likelihood) with regard to singlelevel (Lee & Song, 2004) and multilevel factor models with few clusters (Asparouhov & Muthén, 2010b; Hox, van de Schoot, & Matthijsse, 2012). The possibility to incorporate informative prior information in the estimation process might further increase the applicability of the aforementioned models in small samples (Depaoli & Clifton, 2015; Holtmann, Koch, Lochner, & Eid, 2016; Lee, Song, & Cai, 2010). Bayesian methods may also improve convergence problems in multilevel SEM associated with improper solutions and inadmissible parameter estimates, such as negative variances, by assigning zero prior probability to these parameter spaces (Depaoli & Clifton, 2015; Hox et al., 2012).

Second, Bayesian methods allow researchers to compute credibility intervals for key quantities in longitudinal CFA-MTMM models, for example, with respect to coefficients of convergent and discriminant validity, method specificities, or the stability of states or method effects. In contrast, classical calculations of confidence intervals based on normal theory may be unreliable for these types of parameters (e.g., variance and covariances).

Third, Bayesian methods allow researchers to include information from previous studies in a meta-analytical way. This option seems especially interesting for (longitudinal) MTMM analysis if researchers aim to include past findings concerning the convergent and discriminant validity of a particular measure or instrument in future studies.

Fourth, longitudinal data often suffer from attrition and the problem of missing data. In the presence of missing data that could be missing at random, full information estimation methods should be applied. Bayesian estimation provides a valid full-information approach that can ensure that missing data is properly accounted for in cases where it is computationally not feasible to use maximum likelihood due to the need of numerical integration over high-dimensional integrals (Asparouhov & Muthén, 2010b).

Although Bayesian methods have been successfully applied in IRT modeling (Béguin & Glas, 2001; Fox & Glas, 2001; Hojtink & Molenaar, 1997; Miyazaki & Hoshino, 2009; Patz & Junker, 1999) and longitudinal data analysis (Dunson, 2003; Song, Lu, Hser, & Lee, 2011), they have not been used for the analysis of longitudinal multilevel MTMM data with ordinal response variables. In the next chapters, new models that complement longitudinal CFA-MTMM modeling approaches for the combination of structurally different and interchangeable methods by analyses suitable for ordinal response variables and Bayesian estimation techniques are introduced.

1.4 Basic ideas of the modeling techniques

In the following chapters, different longitudinal multilevel MTMM graded response models for measurement designs combining structurally different and interchangeable methods are introduced: a latent state, a latent change, a latent state-trait, and a latent growth curve GRM. These longitudinal latent variable models belong to the most widely applied CFA approaches to longitudinal data modeling and serve to answer different research questions. Furthermore, all of the models presented in this work are constructively defined based on psychometric theory as well as on design-oriented CFA-MTMM modeling approaches (Eid et al., 2016). In the following, the basic ideas of these modeling approaches are introduced.

1.4.1 Approaches to MTMM modeling

CFA MTMM modeling approaches offer a more accurate estimation of convergent and discriminant validity than could be obtained by correlations of observed variables, separating measurement error from true scores and estimating correlations based on the measurement-error free true score variables.

In design-oriented modeling approaches, model definitions are based on psychometric theory (i.e., stochastic measurement theory) and a well-defined random experiment (Eid et al., 2016). Latent variables are explicitly defined as random variables based on mathematical functions of true score variables, and thereby, have a clear meaning. It has been argued that explicitly taking the measurement design and sampling procedure into account when defining complex CFA models may reduce the risk of estimation and interpretation problems (Eid et al., 2016; Geiser, Koch, & Eid, 2014; Geiser, Bishop, & Lockhart, 2015).

In the context of MTMM modeling, each measured variable is conceived as a trait-method-unit (TMU; Campbell & Fiske, 1959). Each measured variable, that is, each TMU, is assumed to have its own true score variable. As in the basic decomposition of classical psychometric test theory (CTT; Steyer, 1989), each measured variable Y_{jk} can then be decomposed into a true score and an error variable,

$$Y_{jk} = \tau_{jk} + \varepsilon_{jk}, \tag{1.4.1}$$

where Y_{jk} is the observed variable of construct j measured by method k, the true score τ_{jk} represents the expectation of Y_{jk} given a person, i.e., the mean of a person's intraindividual score distribution, and ε_{jk} is an error term.

Latent factors can then be identified by imposing restrictions on the correlation structure of the true score variables. To obtain identified CFA MTMM models in designs with only a single indicator per TMU, it is necessary to impose the assumption that method effects are perfectly correlated across different constructs (Eid, 2000; Marsh & Grayson, 1995). This is a rather strong assumption that is often too restrictive for empirical applications. In order to obtain trait-specific method effects, multiple-indicator MTMM designs have been developed (Eid et al., 2003, 2008; Marsh & Hocevar, 1988; Marsh, 1993). These designs use multiple indicators per TMU and are built on the assumption that true scores are unidimensional across items pertaining to the same TMU (Eid et al., 2003). The most basic multiple-indicator MTMM measurement model building on this assumption is the baseline TMU model by Marsh and Hocevar (1988) depicted in Figure 1.1. The latent correlation matrix between the factors in this model could be interpreted along the same lines as in the

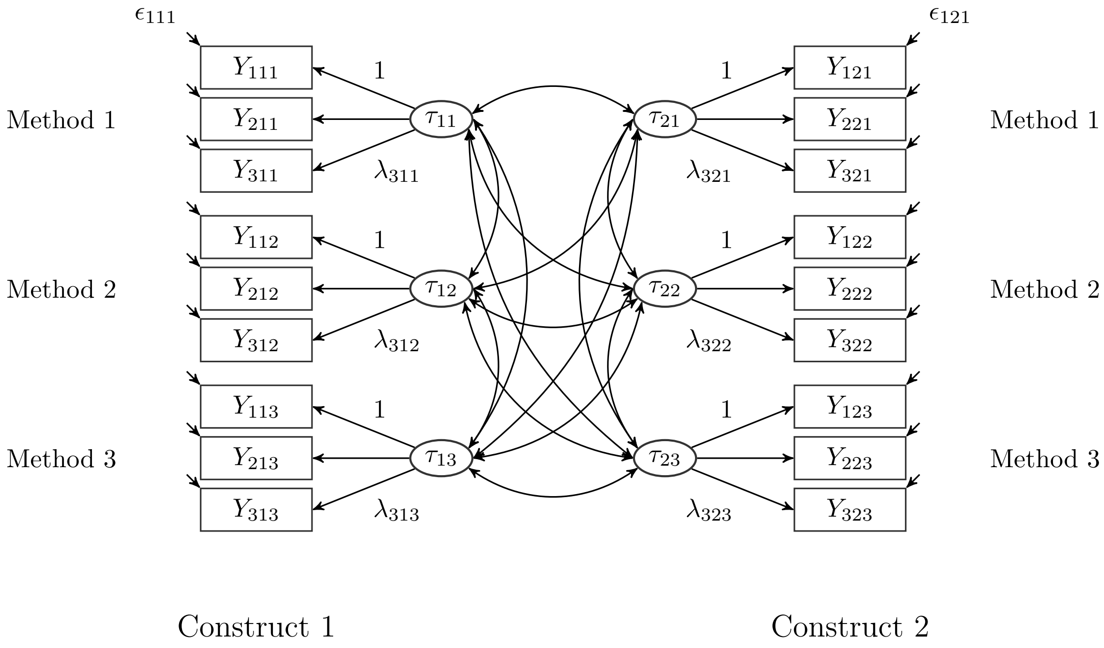

Figure 1.1: Path diagram of the baseline TMU model with two constructs and three (structurally different) methods, measured with three indicators each. ε_{ijk}: residual variable for indicator i of construct j and method k; λ_{ijk}: loading parameter for indicator i of construct j and method k; τ_{jk}: true score variable for construct j and method k; Y_{ijk}: observed variables of the i-th item of construct j measured by method k.

classical MTMM matrix as proposed by Campbell and Fiske (1959), with the difference that the convergent and discriminant validity are being explored at the latent level.

The models presented in this work are based on a design-oriented CFA-MTMM modeling approach (Eid et al., 2016), the CTC(M-1) approach (Eid, 2000; Eid et al., 2003, 2008).

The idea underlying the CTC(M-1) approach is that method effects can be analyzed by defining one method as a reference method and contrasting the remaining methods against this reference. The CTC(M-1) model is thereby an extension of the baseline TMU model, which focuses on different contrasts between the methods. Given certain parameter restrictions, the (restricted) CTC(M-1) model is a perfect reparameterization of the baseline TMU model, i.e., has the same model fit (Geiser, Eid, & Nussbeck, 2008).

In the CTC(M-1) approach, method effects are defined as latent residual variables, reflecting the over- or underestimation of the construct under investigation by the non-reference method with respect to the reference method (Eid et al., 2003, 2008). That is, the true score variables of the non-reference methods are regressed on the true scores of the reference methods, obtaining method effects as residual variables of this latent regression:

$$\mathbb{E}[\tau_{jk} \mid \tau_{j1}] = \alpha_{jk} + \beta_{jk}\tau_{j1} \tag{1.4.2}$$

and

$$M_{jk} := \tau_{jk} - \mathbb{E}[\tau_{jk} \mid \tau_{j1}] \tag{1.4.3}$$

The trait factor in this model corresponds to the true score variable of the reference method

(here τ_{j1}). Consequently, trait and method factors are defined relative to the chosen reference method and there is no method factor for the indicators of the reference method.
It has been argued that the CTC(M-1) model overcomes many methodological problems of the well-known CT-CM model, which has been repeatedly associated with interpretation problems, non-convergence, improper solutions, and vanishing method factors (Eid et al., 2016; Geiser, Bishop, & Lockhart, 2015).
Besides the CTC(M-1) approach, there are other desing-oriented approaches to MTMM modeling that use constructively defined latent variables and are equivalent to the basic TMU measurement model. These are the latent difference (LD) model (Pohl et al., 2008) and latent means (LM) model (Pohl & Steyer, 2010). While in the LD model method factors are defined as latent difference scores with respect to a reference method, the LM model defines the trait factor as the grand mean of the true score variables pertaining to the same construct and method factors as deviations from this average true score variable.
Each of these models has its own strengths and weaknesses. A detailed comparison of the approaches can be found in Geiser, Eid, West, Lischetzke, and Nussbeck (2012) as well as in Koch, Eid, and Lochner (in press). The CTC(M-1) approach was chosen as it features the following advantages. First, it allows researchers to compare methods that are measured on different scales, for instance, contrasting physiological measures with rating scales. In contrast, the LD model requires the items representing different methods to be measured on a common metric (Geiser et al., 2012). Second, Geiser et al. (2012) have shown that the CTC(M-1) model allows to separate three types of method bias: general, conditional, and individual method bias. In contrast, individual and conditional method bias are not separable in the LD and LM models. The conditional method bias in the CTC(M-1) model depends on the deviation of the regression line from a 45 line (Geiser et al., 2012). That is, it depends on the regression intercept and regression slope, the latter of which quantifies the degree to which the over- or underestimation by the non-reference method depends on the value of the reference method. The individual values of the residual and its variance represent the degree of individual bias, that is, the deviation of an individual value from the value predicted by the regression. While the conditional method bias is perfectly correlated with the reference method factor, the residual represents a pure method effect that is corrected for influences of the reference method (Geiser et al., 2012). Third, the CTC(M-1) approach implies an additive decomposition of the observed variables' variances and thereby allows to compute additive variance components associated with trait or method effects (e.g., quantifying the degree of convergent validity). This additive decomposition in the CTC(M-1) model is possible as method effects have been corrected for reference trait influences, while method effects can be correlated with trait influences in the LD and LM models.

1.4.2 Situational influences

As Steyer, Mayer, Geiser, and Cole (2015, p. 71) postulate in their revision of latent state-trait (LST) theory, "observations are fallible, they never happen in a situational vacuum, they are always made using a specific method of observations, and there is no person without a past".
While the necessity to account for measurement error and method-specificity of observations was discussed in the previous sections, the following sections turn to the issue of situational influences and the representation of change in models of longitudinal data analysis.
All of the models presented in this work incorporate the basic idea of LST theory (Steyer,

Ferring, & Schmitt, 1992; Steyer, Schmitt, & Eid, 1999) that psychological measurements do not take place in a situational vacuum. Additionally, they take into account that targets and raters may be affected differently by inner and outer aspects of a particular situation. Although targets and raters may experience the same outer situation (e.g., answering a questionnaire), their inner situations may be quite different (they may, for instance, be happy, tired, or stressed). The term situation will be used in a broader sense in this work, that is, encompassing all inner and outer aspects of a situation that can affect a person (target or rater) at a specific time and place. Note that this definition of a situation is not restricted to the experimental situation in the moment of the assessment (measurement, rating). Instead, it encompasses all experiences and circumstances that might temporarily affect a person's measurement (rating), regardless of whether they are a characteristic of the measurement situation per se (e.g., the temperature in the room where the assessment takes place) or not (e.g., a stressful day at work influencing the target's or rater's present mood).

1.4.3 Longitudinal CFA modeling

The longitudinal latent variable models treated in this work (latent state, latent change, latent state-trait and latent growth curve) belong to the most widely applied CFA approaches to (continuous) longitudinal data modeling (Newsom, 2015). Note that the models in this work address longitudinal modeling of continuous latent variables. That is, approaches for categorical latent variables, such as latent class or latent transition analysis (Collins & Flaherty, 2002; Eid, 2006; Langeheine, 1994), as well as approaches aiming at predicting the (time of) occurrence of a discrete event, such as survival analysis (Cox & Oakes, 1984), are not discussed here.

Khoo, West, Wu, and Kwok (2006) identify three general classes of longitudinal models for analyzing change: autoregressive models, growth curve models, and latent state-trait models. Similarly, Eid and Langeheine (1999) distinguish between five groups of models, depending on the kind of change they assume: latent state models, autoregressive models, growth curve models, variability models, and autoregressive variability models. Latent state (LS) models represent the simplest models for analyzing change. In LS models, the covariance structure of the latent variables is not restricted, and change is quantified indirectly via the strength of associations between the latent variables over time. Different restrictions of the covariance structure between the latent variables in LS models lead to one of the more complex models. Collins (2006) argues that the first step to longitudinal data analysis should always be to consider the theoretical model of change, that is, the nature of the change phenomenon that is to be studied. In this framework, variability and change are treated as distinct processes that can be distinguished (Eid & Kutscher, 2014; Nesselroade & Ram, 2004). While variability refers to short-term, reversible fluctuations around a stable set-point (time-constant trait), change is conceptualized to be more enduring and independent of short-term influences (i.e., trait change; Nesselroade, 1991; Eid & Kutscher, 2014; Geiser, Keller, et al., 2015; Steyer et al., 2015).

Hence, the different models obtained by imposing restrictions on the latent association structure over time assume different processes of change and serve to answer different research questions. While LS models can be used to investigate the stability of constructs via their correlations over time, latent change (LC) models present a reparameterization of LS models, parameterizing change as latent difference scores of the latent states between measurement occasions. Variances of the latent difference score factors then serve to quantify the amount of inter-individual differences in intra-individual change. Latent first-order autoregressive models, in contrast, model longitudinal processes by regressing a latent variable on the latent variable at the preceding time point. Change is then defined as the regression residual of this latent autoregression. Autoregressive models implicitly assume that stability declines with an increasing time lag and that the process is characterized by temporal inertia (Hertzog & Nesselroade, 1987). Consequently, these models are typically applied to model long-lasting and irreversible change (Eid & Langeheine, 1999). Similar to LC models, latent growth curve (LGC) models (Bollen & Curran, 2006; McArdle & Epstein, 1987) serve to investigate inter-individual differences in intra-individual change. However, LGC models impose the restriction that change follows a function of time, thereby allowing to estimate person-specific growth curves and identify trends in the data. In contrast, latent state-trait (LST) models (Steyer et al., 1992, 1999) are commonly understood as models focusing on variability. They allow to evaluate the variability of a construct by differentiating between occasion-specific and stable influences.

Change is a phenomenon that occurs within the individual, making intraindividual variability the primary interest in modeling longitudinal data (Collins, 2006; Hertzog & Nesselroade, 1987). Accordingly, old as well as newer approaches stress the importance of separating inter- and intraindividual variability, i.e., the between-person from the within-person level, in the specification of lagged-panel or autoregressive models (e.g., Hamaker, Kuiper, & Grasman, 2015; Hertzog & Nesselroade, 1987; Rovine & Walls, 2006). While Rovine and Walls (2006) have introduced an autoregressive model in the context of multilevel analysis, Hamaker et al. (2015) focused on the cross-lagged panel model and argued that it is essential to model stable inter-individual differences by the inclusion of a latent trait or random intercept in the model. By inclusion of a stable latent trait factor in autoregressive panel models, the modeling approach essentially becomes equivalent to models separating trait and occasion-specific components of an observation (i.e., LST models) with autoregressive effects on the occasion-specific components (Cole, Martin, & Steiger, 2005; Luhmann, Schimmack, & Eid, 2011). One difference is that the model by Hamaker et al. (2015) does not take measurement error into account. A multiple-indicator, latent-variable variant of this model, termed the Trait-State-Occasion (TSO) model, was introduced by Cole et al. (2005), who argued that autoregressive effects should be modeled on the level of the occasion-specific (residual) variables instead of the latent state variables to avoid a recursiveness in the model. Other approaches have integrated autoregressive structures in latent growth models (e.g., Bollen & Curran, 2004). These models fall into the category of "autoregressive variability models" in the classification given by Eid and Langeheine (1999). The class of autoregressive variability models comprises those models that assume a combination of different kinds of change, such as incorporating latent time-constant and occasion-specific variables with an autoregressive structure into a hybrid model (Eid & Langeheine, 1999). Due to the aforementioned shortcoming of failing to differentiate between inter- and intraindividual levels of change, the classical autoregressive model is not among the models covered in this work. Instead, autoregressive models will be merely discussed in the context of LST models in the following.

1.5 Aims and scope of the present work

The aim of the present work is to present longitudinal multilevel CFA-MTMM models for ordered categorical response variables. The presented models combine the modeling possibilities of the continuous indicator models for measurement designs combining structurally different and interchangeable methods introduced by Koch (2013) with the advantages of an IRT approach to analyzing longitudinal MTMM data. In addition, it will be shown how Bayesian estimation techniques can address a number of important issues that typically arise in longitudinal multilevel MTMM studies. Bayesian estimation techniques do not only render the estimation of complex MTMM-IRT models possible, but also offer additional advantages over classical frequentist approaches. The performance of the models is investigated in three Monte Carlo simulation studies. Additionally, the practical use of the models is illustrated with regard to an empirical application to subjective well-being data. Advantages and limitations of the presented models and their estimation using Bayesian methods are discussed.

All of the models presented in this work build on the definition of latent variables on a specified random experiment. This random experiment and the probability space used to define the random variables are introduced within the definition of the LS-Com GRM (Chapter 2), however, apply to all of the following models, too.

Chapter 2

Latent State (LS-Com) Graded Response Model

2.1 General introduction to the model definitions

In the following chapters, different longitudinal multilevel MTMM graded response models for measurement designs combining structurally different and interchangeable methods are introduced. The graded response model (Samejima, 1969, 2010) was chosen as it is equivalent to the measurement model of CFA models for response variables with ordered response categories (Takane & De Leeuw, 1987). As models for measurement designs combining structurally different and interchangeable methods, the presented models are multilevel models. Recall that interchangeable methods are those methods that are randomly drawn from a set of equivalent methods, such as multiple peer ratings for the same target. As they are the result of a multi-stage sampling procedure, interchangeable raters are measured on the within-level. It is assumed that the interchangeable raters are fully nested within targets (i.e., there is no cross-classification structure). An adequate approach to model these data has to take the resulting multilevel structure into account (e.g., Eid et al., 2008; Koch et al., 2014). Ignoring the dependencies in multilevel structures can have detrimental effects such as biased parameter estimates and standard errors (Julian, 2001).
Structurally different methods, in contrast, are methods that cannot be easily replaced by one another, such as self-ratings or a parent rating of a child's characteristics. In contrast to the interchangeable raters, the structurally different raters are fixed given the target. Hence, there is no additional sampling step for the structurally different raters once a target has been selected. Therefore they are measured on the target-level. However, also the responses of the structurally different raters are observed in rater-specific situations.
As longitudinal MTMM models, the presented GRMs are designed to model data including measures on multiple constructs j that are assessed with different methods k on multiple measurement occasions l. Furthermore, each construct is assumed to be measured by several indicators i (for a list of indices used in the model definition see Table 2.1). The observed variables are assumed to be measured on a rating scale with ordered categorical response options $s \in S_{ij}$, with $S_{ij} = \{0, \cdots, q_{ij} - 1\}$, where q_{ij} is the number of response categories of item i of construct j.
The following model definitions are based on a model with one set of interchangeable and

Table 2.1: List of Indices used in the Model Definitions.

Index	Meaning
r	rater
t	target
i	indicator
j	construct
k	method
l	measurement occasion
s	response category
q_{ij}	number of response categories of indicator i of construct j

two structurally different methods. The index k serves for the distinction between reference and non-reference as well as structurally different and interchangeable methods. Choosing the first, structurally different method ($k = 1$) as the reference method, the second method ($k = 2$) indicates the non-reference interchangeable method (e.g., peer reports). All methods $k > 2$ are assumed to be structurally different non-reference methods. Adding an additional set of interchangeable or structurally different methods is straightforward. The basic structure of the models will be explained using raters as methods (Kenny, 1995). Thus, a self-report (structurally different reference method), different interchangeable peer ratings (interchangeable non-reference method), and parent reports (structurally different non-reference method) serve as an example.

2.2 The random experiment, probability space, and conditional probability distributions

First, the random experiment that characterizes the sampling procedure for longitudinal measurement designs of a graded response model with structurally different and interchangeable methods is specified[1]. Based on the specified random experiment, the variables in the LS-Com, LC-Com, LST-Com, and LGC-Com GRM can then be properly defined as random variables. This approach is based on stochastic measurement theory following the approach by Eid (2000), Koch, Eid, and Lochner (in press), Steyer and Eid (2001) as well as Steyer (1988). The latent variables in the GRMs defined in the following chapters are thereby random variables that are well-defined on a specified random experiment and, therefore, have a clear meaning. Note that recently Steyer et al. (2015) proposed a revision of LST (LST-R) theory building on a different definition of the random experiment, in which the probability space Ω explicitly includes persons' experiences between measurement occasions. It is noteworthy that, although the conceptualization of the random experiment and the definition of the probability space given in the following differs from that in LST-R theory, the vast majority of models that can be specified on the basis of these theories are identical. That is, all of the models defined in the following sections could also be defined based on LST-R theory. LST-R theory is discussed in Section 4.13.
The following model definitions are designed for data including measures on multiple constructs j that are assessed with different methods k using several indicators i on multiple

[1]Note that parts of the this chapter have been pre-published in Holtmann et al. (2017)

measurement occasions l. Let the first method $k = 1$ be a structurally different self-report method, the second method ($k = 2$) an interchangeable method, and all methods $k > 2$ structurally different informant report methods. Then, consider the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with the set

$$\Omega = \Omega_T \times \Omega_{TS_1} \times \ldots \times \Omega_{TS_l} \times \ldots \times \Omega_{TS_f} \times \Omega_R \times \Omega_{R_2S_1} \times \ldots \times \Omega_{R_kS_l} \times \ldots \times \Omega_{R_eS_f} \times \Omega_O$$

where Ω_T is the set of targets assessed within a target-specific situation Ω_{TS_l} on measurement occasion $l \in L = \{1, \cdots, f\}$, Ω_R is the set of interchangeable raters, and $\Omega_{R_kS_l}$ is the set of possible rater-situations for rater k, $k > 1$, on measurement occasion l. Note that the set of interchangeable raters Ω_R does not get an index (k = 2) for simplicity reasons, as the random experiment is defined for one set of interchangeable raters only. Ω_O is the set of possible outcomes given by

$$\Omega_O = \Omega_{O_{11}} \times \cdots \times \Omega_{O_{kl}} \times \cdots \times \Omega_{O_{ef}} \tag{2.2.1}$$

where each $\Omega_{O_{kl}}$ for method $k \in K = \{1, \cdots, e\}$ and measurement occasion l is the product set $\Omega_{O_{kl}} = A_{1kl} \times \cdots \times A_{jkl} \times \cdots \times A_{dkl}$. Each set A_{jkl} is the cross-product of the c_j sets O_{ijkl}, $A_{jkl} = O_{1jkl} \times \cdots \times O_{ijkl} \times \cdots \times O_{c_jjkl}$, containing the possible outcomes for item i, $i \in I_j = \{1, \cdots, c_j\}$, of construct j, method k and occasion l.
Define the projections $p_T : \Omega \rightarrow \Omega_T$ as the mapping of the possible outcomes to the set of targets, the projection $p_R : \Omega \rightarrow \Omega_R$ as the mapping of the possible outcomes to the set of interchangeable raters, the projection $p_{TS_l} : \Omega \rightarrow \Omega_{TS_l}$ as the mapping of the possible outcomes to the set of target-situations, and the projection $p_{R_kS_l} : \Omega \rightarrow \Omega_{R_kS_l}$ as the mapping of the possible outcomes to the set of rater-situations for the raters of method k.
The variables Y_{rtij2l} and Y_{tijkl} are random variables on $(\Omega, \mathcal{A}, \mathcal{P})$, defined by the following mappings: (1) for a level-1 observation belonging to an interchangeable rater the variable Y_{rtij2l} is defined as $Y_{rtij2l} : \Omega_T \times \Omega_{TS_l} \times \Omega_R \times \Omega_{R_2S_l} \times \Omega_O \rightarrow S_{ij}$, (2) for a level-2 observation as rated by a structurally different rater other than the target, $Y_{tijkl} : \Omega_T \times \Omega_{TS_l} \times \Omega_{R_kS_l} \times \Omega_O \rightarrow S_{ij}$, with $k > 2$, and (3) for a level-2 self-report observation (structurally different), the observed variable is defined by the projection $Y_{tij1l} : \Omega_T \times \Omega_{TS_l} \times \Omega_O \rightarrow S_{ij}$, with $S_{ij} = \{0, \cdots, q_{ij} - 1\}$, where q_{ij} is the number of response categories of item i of construct j (cf. Eid, 1995; Koch, 2013). Note that, for the sake of simplicity, the possible response categories for an item i of construct j are assumed to be equal across methods and measurement occasions.
The observed values of an indicator i of construct j, assessed by an interchangeable method $k = 2$, on the l^{th} occasion of measurement for target t as rated by rater r are denoted by the values of the level-1 variable Y_{rtij2l}. These variables are measured on the rater-level (Level 1) as they are not only target- but also rater-specific. In contrast, the variables Y_{tijkl}, $k \neq 2$, are level-2 variables, and the values of Y_{tijkl} are the observed values of indicator i for construct j, assessed by a structurally different method k on the l^{th} occasion of measurement for target t. Only the interchangeable methods are measured on the within-level, as they are the result of a multi-stage sampling procedure. In contrast, the structurally different raters are fixed given the target. Hence, there is no additional sampling step for the structurally different raters once a target has been selected. However, also the responses of the structurally different raters are observed in rater-specific situations, and thus depend on the projection $p_{R_kS_l}$. Note that the model definitions of the GRMs given here deviate from their continuous

indicator model counterparts in that the latter did not contain rater-specific situations for the structurally different raters and hence did not include the projection $p_{R_k S_l}$ for $k > 2$.
The set of rater situations on measurement occasion l is assumed to include all the situations any of the raters could potentially encounter. As stated earlier, the term situation refers to all inner and outer situations encompassing all experiences and circumstances that might temporarily affect a person's measurement (rating) at a specific time and place. In order to distinguish between the rater-situations of the structurally different and interchangeable raters, the subscript k is used.
The graded response model in normal ogive form is based on cumulative probit link models (Agresti, 2007; B. Muthén & Asparouhov, 2002). For the present model, latent response variables $\pi_{rtsijkl}$ and π_{tsijkl} are defined for each observed variable Y_{rtijkl} and Y_{tijkl}, respectively, and for each category $s \in S_{ij}$ by the application of a probit link (Eid, 1995, 1996):

$$\pi_{tsij1l} := \Phi^{-1}[P(Y_{tij1l} \geq s \mid p_T, p_{TS_l})] \tag{2.2.2}$$
$$\pi_{rtsij2l} := \Phi^{-1}[P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l})] \tag{2.2.3}$$
$$\pi_{tsijkl} := \Phi^{-1}[P(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_k S_l})] \qquad k > 2 \tag{2.2.4}$$

with Φ denoting the cumulative distribution function of the standard normal distribution, taking $k = 1$ as the reference method, $k = 2$ as the interchangeable non-reference method, $k > 2$ as structurally different non-self-report methods, and

$$P(Y_{tij1l} \geq s \mid p_T, p_{TS_l}) = \mathbb{E}[I_{\{Y_{tij1l} \geq s\}} \mid p_T, p_{TS_l}] \tag{2.2.5}$$
$$P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l}) = \mathbb{E}[I_{\{Y_{rtij2l} \geq s\}} \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l}] \tag{2.2.6}$$
$$P(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_k S_l}) = \mathbb{E}[I_{\{Y_{tijkl} \geq s\}} \mid p_T, p_{TS_l}, p_{R_k S_l}] \qquad k > 2, \tag{2.2.7}$$

where $\mathbb{E}[\cdot \mid \cdot]$ denotes the conditional expectation and I denotes the indicator function with, e.g., $I_{\{Y_{tij1l} \geq s\}} = 1$, if $Y_{tij1l} \geq s$, and $I_{\{Y_{tij1l} \geq s\}} = 0$, otherwise. It is assumed that common latent response variables π_{tijkl} and π_{rtij2l} exist, which are common latent variables of all latent variables π_{tsijkl} and $\pi_{rtsij2l}$, respectively, belonging to the same item i of construct j at time l:

$$\pi_{tijkl} := \pi_{tsijkl} + \kappa_{sijkl} \qquad k \neq 2 \tag{2.2.8}$$
$$\pi_{rtij2l} := \pi_{rtsij2l} + \kappa_{sij2l} \tag{2.2.9}$$

The constants κ_{sijkl} and κ_{sij2l} represent the position of the normal ogives on the latent continua π_{tijkl} and π_{rtij2l}, their values being the inflection points of the normal ogives that depict the dependence of the probabilities $P(Y_{tij1l} \geq s \mid p_T, p_{TS_l})$, $P(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_k S_l})$, $k > 2$, and $P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l})$ on the latent response variables π_{tij1l}, π_{tijkl}, $k > 2$, and π_{rtij2l}, respectively. They can be interpreted as difficulty parameters of the dichotomized variables. The probability functions $P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l})$ and $P(Y_{tij1l} \geq s \mid p_T, p_{TS_l})$ can thus be expressed as:

$$\begin{aligned} P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2 S_l}) &= P(Y_{rtij2l} \geq s \mid \pi_{rtij2l}) \\ &= \Phi(\pi_{rtij2l} - \kappa_{sij2l}) = \int_{-\infty}^{\pi_{rtij2l} - \kappa_{sij2l}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \end{aligned} \tag{2.2.10}$$

$$
\begin{aligned}
P(Y_{tij1l} \geq s \mid p_T, p_{TS_l}) &= P(Y_{tij1l} \geq s \mid \pi_{tij1l}) \\
&= \Phi(\pi_{tij1l} - \kappa_{sij1l}) = \int_{-\infty}^{\pi_{tij1l}-\kappa_{sij1l}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx
\end{aligned}
\tag{2.2.11}
$$

and analogously for Y_{tijkl}, $k > 2$.
The higher a person's score on the latent response variable, the higher the probability that this person in a certain situation will respond in category s or higher to the respective item. The conditional probability that Y_{rtij2l} takes on the value of category s is then given by (Eid, 1995):

$$
\begin{aligned}
P(Y_{rtij2l} = s \mid \pi_{rtij2l}) &= P(Y_{rtij2l} \geq s \mid \pi_{rtij2l}) - P(Y_{rtij2l} \geq s+1 \mid \pi_{rtij2l}) \\
&= \Phi(\pi_{rtij2l} - \kappa_{sij2l}) - \Phi(\pi_{rtij2l} - \kappa_{(s+1)ij2l})
\end{aligned}
\tag{2.2.12}
$$

with $P(Y_{rtij2l} \geq 0 \mid \pi_{rtij2l}) = 1$ and $P(Y_{rtij2l} \geq s \mid \pi_{rtij2l}) = 0 \;\text{ for }\; s > q_{ij} - 1$. The conditional probabilities $P(Y_{tijkl} = s \mid \pi_{tijkl})$, $k \neq 2$, are computed in the same manner.

2.3 Introduction to the LS-Com GRM

This chapter introduces the Latent-State-Combination-Of-Methods-Graded-Response-Model (LS-Com GRM). The model is based on the LS-Com model for continuous indicator variables developed by Koch (2013); Koch et al. (2014). The present model combines the modeling possibilities of the continuous indicator LS-Com model with the advantages of an IRT approach to analyzing longitudinal MTMM data. The LS-Com GRM allows to

1. analyze convergent and discriminant validity over time,
2. specify method factors on different measurement levels,
3. analyze change and stability of construct and method effects over time,
4. investigate the generalizability of method effects across methods or time,
5. test the degree of measurement invariance over time on the item-level,
6. compare item difficulties, item discrimination and reliability for different methods, and
7. test mean changes of constructs over time.

The following derivation of the latent variables in the LS-Com GRM build on the definition of the random experiment and the latent response variables in Section 2.2.
According to these definitions, the latent response variables π_{tijkl}, $k \neq 2$, are target specific, measured on Level 2, while the latent response variables π_{rtij2l} are rater-target specific, and thereby measured on Level 1. Having defined the latent response variables π_{tijkl} and π_{rtij2l}, the latent state and method variables can be defined in terms of conditional expectations (analogous to the continuous indicator case; Koch, 2013). Selecting the second method ($k = 2$) as the set of interchangeable (non-reference) methods, each latent variable π_{rtij2l} can be decomposed into the following random variables on $(\Omega, \mathcal{A}, \mathcal{P})$:

$$
S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \tag{2.3.1}
$$

$$\begin{aligned} UM_{rtij2l} &= \pi_{rtij2l} - \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \\ &= \pi_{rtij2l} - S_{tij2l} \end{aligned} \tag{2.3.2}$$

The target-level latent state variables S_{tij2l} can be conceived as the expected peer rating of target t across the interchangeable peer ratings for this target on occasion l for indicator i measuring construct j. That is, they can be considered as the occasion-specific "true means" of the interchangeable ratings for this target and are thereby rater-unspecific. The unique method variables UM_{rtij2l} are rater-specific level-1 variables. They represent the true occasion-specific unique deviation of a particular rater from the expected rating over all interchangeable raters for target t (S_{tij2l}), that is, the over- or underestimation of the true expected peer rating by a particular rater r (Koch, 2013). Due to their definition as latent residual variables, the UM_{rtij2l} variables have an expectation of zero and are uncorrelated with the level-2 latent state variables S_{tij2l}.
The latent state variable of the structurally different self-report S_{tij1l} as well as structurally different informant reports S_{tijkl}, $k > 2$, measured on Level 2, correspond to the common latent variables π_{tijkl} for indicator i, construct j and time point l:

$$S_{tijkl} = \pi_{tijkl} \qquad k \neq 2 \tag{2.3.3}$$

Having defined the target-level latent variables S_{tij2l} for the interchangeable method, the latent state variables belonging to different types of methods (S_{tij2l} and S_{tijkl}) are measured on the same level (Level 2) and can be contrasted against each other. This idea follows the CTC($M-1$) approach for structurally different methods (Eid, 2000; Eid et al., 2003, 2008), regressing the non-reference latent state variables on the latent state variables of the reference-method:

$$\mathbb{E}[S_{tijkl} \mid S_{tij1l}] = \lambda_{Sijkl} S_{tij1l} \qquad k \neq 1 \tag{2.3.4}$$

Note that the regression equations do not include intercepts, as intercept and threshold parameters κ_{sijkl} are not separately identifiable (see Sections 2.5 and 2.13). The residuals of this latent regression analysis can be defined as latent method variables on the target-level, the common method variables CM_{tij2l} and the method variables M_{tijkl}, $k > 2$:

$$CM_{tij2l} = S_{tij2l} - \mathbb{E}[S_{tij2l} \mid S_{tij1l}] \tag{2.3.5}$$

$$M_{tijkl} = S_{tijkl} - \mathbb{E}[S_{tijkl} \mid S_{tij1l}] \qquad \forall \; k > 2 \tag{2.3.6}$$

The CM_{tij2l} variables are that part of the true expected rating of the interchangeable raters (e.g. peer ratings) that is not shared with the reference method (here: self-ratings). The latent variables CM_{tij2l} are termed common method variables, as they represent a common view of the interchangeable raters on the target, that is not shared with the self-reports (on a particular occasion of measurement for construct j; Koch, 2013). The method variables M_{tijkl} for structurally different methods $k > 2$ represent the part of the true informant rating that is not shared with the reference method. Due to their definition as latent residual variables, the common method variables as well as the method variables are uncorrelated with the latent state variables of the reference method (S_{tij1l}) and have an expectation of zero (Koch, 2013). Assuming that the latent method variables of the same method differ only by multiplicative constants, common latent method factors can be defined (Koch, 2013):

$$CM_{tij2l} = \lambda_{CMij2l} CM_{tj2l} \tag{2.3.7}$$

$$UM_{rtij2l} = \lambda_{UMij2l} UM_{rtj2l} \tag{2.3.8}$$

$$M_{tijkl} = \lambda_{Mijkl} M_{tjkl} \qquad k > 2 \tag{2.3.9}$$

This assumption implies that the method variables of different indicators i but belonging to the same construct j and measurement occasion l are perfectly correlated. Note that it is necessary to make assumptions (2.3.8) and (2.3.9) for identifiability reasons, as a model with indicator-specific factors UM_{rtij2l} or M_{tijkl}, $k > 2$, would not be identified.

Overall, the common latent response variables of the non-reference interchangeable method ($k = 2$) in the LS-Com GRM can be expressed as:

$$\begin{aligned} \pi_{rtij2l} &= S_{tij2l} + \lambda_{UMij2l} UM_{rtj2l} \\ &= \lambda_{Sij2l} S_{tij1l} + \lambda_{CMij2l} CM_{tj2l} + \lambda_{UMij2l} UM_{rtj2l} \end{aligned} \tag{2.3.10}$$

The measurement equation for the non-reference structurally different methods is given by:

$$\pi_{tijkl} = \lambda_{Sijkl} S_{tij1l} + \lambda_{Mijkl} M_{tjkl} \qquad k > 2 \tag{2.3.11}$$

Additionally, it could be assumed that the indicator-specific latent state variables are perfectly correlated, resulting in one common latent state factor per construct j and time point l:

$$\pi_{tij1l} = S_{tij1l} = \lambda_{Sij1l} S_{tj1l} \tag{2.3.12}$$

The LS-Com GRM for one interchangeable and two structurally different methods with indicator-specific latent state factors is depicted in Figure 2.1, the analogous LS-Com GRM with common latent state factors is depicted in Figure 2.2.

Based on the above definition of the latent state and latent method variables, the variance of the latent response variables π_{tijkl} and π_{rtij2l} can be additively decomposed into different variance components. The additive decomposition is based on the fact that all of the method factors are defined as latent residual variables with regard to their respective latent state variables and are, therefore, uncorrelated with their regressors (see Section 2.10). The variances of the latent response variables can hence be additively decomposed in the following ways:

$$Var(\pi_{tij1l}) = Var(S_{tij1l}) \tag{2.3.13}$$

$$Var(\pi_{rtij2l}) = \lambda^2_{Sij2l} Var(S_{tij1l}) + \lambda^2_{CMij2l} Var(CM_{tj2l}) + \quad \lambda^2_{UMij2l} Var(UM_{rtj2l}) \tag{2.3.14}$$

$$Var(\pi_{tijkl}) = \lambda^2_{Sijkl} Var(S_{tij1l}) + \lambda^2_{Mijkl} Var(M_{tjkl}) \qquad k > 2 \tag{2.3.15}$$

Analogous to the LS-Com model with continuous indicators, different variance components for the non-reference method indicators can be defined. The following variance components correspond to the coefficients introduced by Koch (2013), with the only difference that they are defined based on the latent response variables π_{tijkl} and π_{rtij2l}.

The consistency coefficients are defined as:

$$Con(\pi_{tijkl}) = \frac{\lambda^2_{S_{ijkl}} Var(S_{tij1l})}{\lambda^2_{S_{ijkl}} Var(S_{tij1l}) + \lambda^2_{M_{ijkl}} Var(M_{tjkl})} \qquad k > 2 \tag{2.3.16}$$

$$Con(\pi_{rtij2l}) = \frac{\lambda^2_{S_{ij2l}} Var(S_{tij1l})}{\lambda^2_{S_{ij2l}} Var(S_{tij1l}) + \lambda^2_{CM_{ij2l}} Var(CM_{tj2l}) + \lambda^2_{UM_{ij2l}} Var(UM_{rtj2l})} \tag{2.3.17}$$

They represent that part of the non-reference method's latent response variable's variance that can be explained by the reference method. For structurally different non-reference methods ($k > 2$), the consistency coefficient is the amount of true interindividual differences in the informant report that is shared with the reference method. In case of interchangeable non-reference methods, the consistency coefficient represents the part of the true interindividual differences between the individual informant reports on the rater-level that is shared with the reference method. The square root of the consistency coefficient can be interpreted as an indicator of the convergent validity between the reference and the non-reference method. Furthermore, for the interchangeable non-reference method an additional consistency coefficient on the target-level can be defined,

$$Con(\pi_{tij2l}) = \frac{\lambda^2_{S_{ij2l}} Var(S_{tij1l})}{\lambda^2_{S_{ij2l}} Var(S_{tij1l}) + \lambda^2_{CM_{ij2l}} Var(CM_{tj2l})} , \tag{2.3.18}$$

representing the amount of interindividual differences in the expected rating over all interchangeable raters that can be explained by the reference method (Koch et al., 2014). Furthermore, three different method specificity coefficients can be defined. The unique method specificity coefficient

$$UMS(\pi_{rtij2l}) = \frac{\lambda^2_{UM_{ij2l}} Var(UM_{j2l})}{Var(\pi_{rtij2l})} \tag{2.3.19}$$

that quantifies the proportion of the variance of an interchangeable method variable that is due to the unique views of the raters, neither shared with the reference-method nor with the other interchangeable raters. In contrast, the common method specificity coefficient indicates the degree of interindividual differences in the interchangeable raters' ratings that goes back to a common view of the raters but is not shared with the reference method (Koch, 2013; Eid et al., 2008):

$$CMS(\pi_{rtij2l}) = \frac{\lambda^2_{CM_{ij2l}} Var(CM_{j2l})}{Var(\pi_{rtij2l})} \tag{2.3.20}$$

The method specificity coefficient $MS(\pi_{tijkl})$ for the structurally different non-reference method represents the amount of the indicator's latent response variable's variance that is not shared with the reference method:

$$MS(\pi_{tijkl}) = \frac{\lambda^2_{M_{ijkl}} Var(M_{jkl})}{Var(\pi_{tijkl})} \qquad k > 2 \tag{2.3.21}$$

In analogy to the continuous indicator LS-Com model, indicator reliability coefficients can also be computed in the LS-Com GRM. Furthermore, a true (i.e., measurement-error free)

intraclass correlation coefficient (ICC) can be computed, indicating how much of the reliable variance of the within-level variables is accounted for by between-level inter-individual differences. See Section 2.11 for definitions.
While convergent validity in the LS-Com GRM is quantified by the consistency coefficients, discriminant validity is reflected in the correlation coefficients between the latent state variables of different constructs j (high correlations indicating low discriminant validity). Furthermore, correlations between method variables of the same type but different constructs j are a measure of the generalizability of method effects across constructs. Correlations between latent state variables or method variables of different measurement occasions l, on the other hand, can be interpreted in terms of construct or method effect stability, respectively. For an exhaustive list of interpretations of the permissible correlations between factors in the LS-Com model see Koch (2013). Note that the latent state variables are uncorrelated with any method variable of the same construct j and measurement occasion l by definition. Also, the unique method variables are uncorrelated with the level-2 latent method variables (see Section 2.10 for details).
In the following sections, the LS-Com GRM will be formally defined. The uniqueness of the latent variables and their coefficients, as well as admissible transformations and meaningful statements regarding the former are discussed. Necessary independence assumptions are introduced and testable consequences for the covariance structure of the model are derived. Last but not least, identification conditions for the model are presented and conditions for testing measurement invariance in the LS-Com GRM are discussed.
Note that the LS-Com GRM is an extension of the LS-Com model for continuous indicators (Koch, 2013). As such, some of the properties of the LS-Com continuous indicator model apply to the LS-Com GRM with no or only minor modifications. These are the existence, uniqueness, admissible transformations and meaningful statements concerning the latent method variables (and their coefficients). These properties are elaborated and proofed in detail by Koch (2013) and will only be reported shortly in Sections 2.4 and 2.5. All other properties deviate from those of the continuous indicator case and are defined in detail in the following sections.

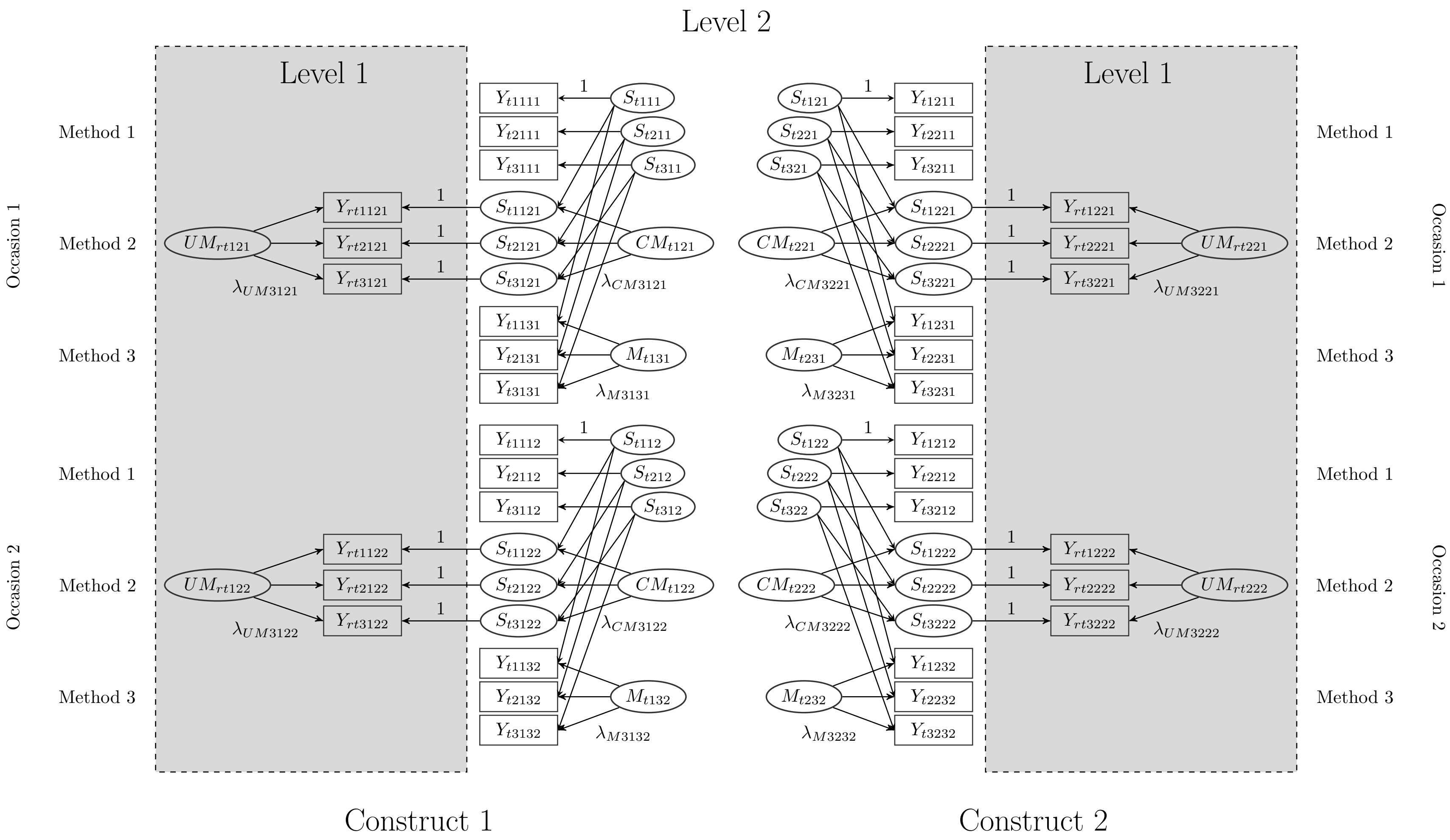

Figure 2.1: Path diagram of the Latent-State-Com graded response model with indicator-specific latent state variables. The model is depicted for two structurally different methods and one set of interchangeable methods at two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For convenience, the constant indicator $k = 1$ has been dropped from the latent state variables ($S_{tijl} = S_{tij1l}$). For the sake of clarity, correlations between latent variables are omitted and loading parameters are only shown for exemplary indicators. Correlations that are not permissible by definition of the LS-Com GRM are correlations between the latent state variables and the latent (common) methods variables of the same construct j and occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method variable; M: method variable; S: latent state variable; UM: unique method variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l.

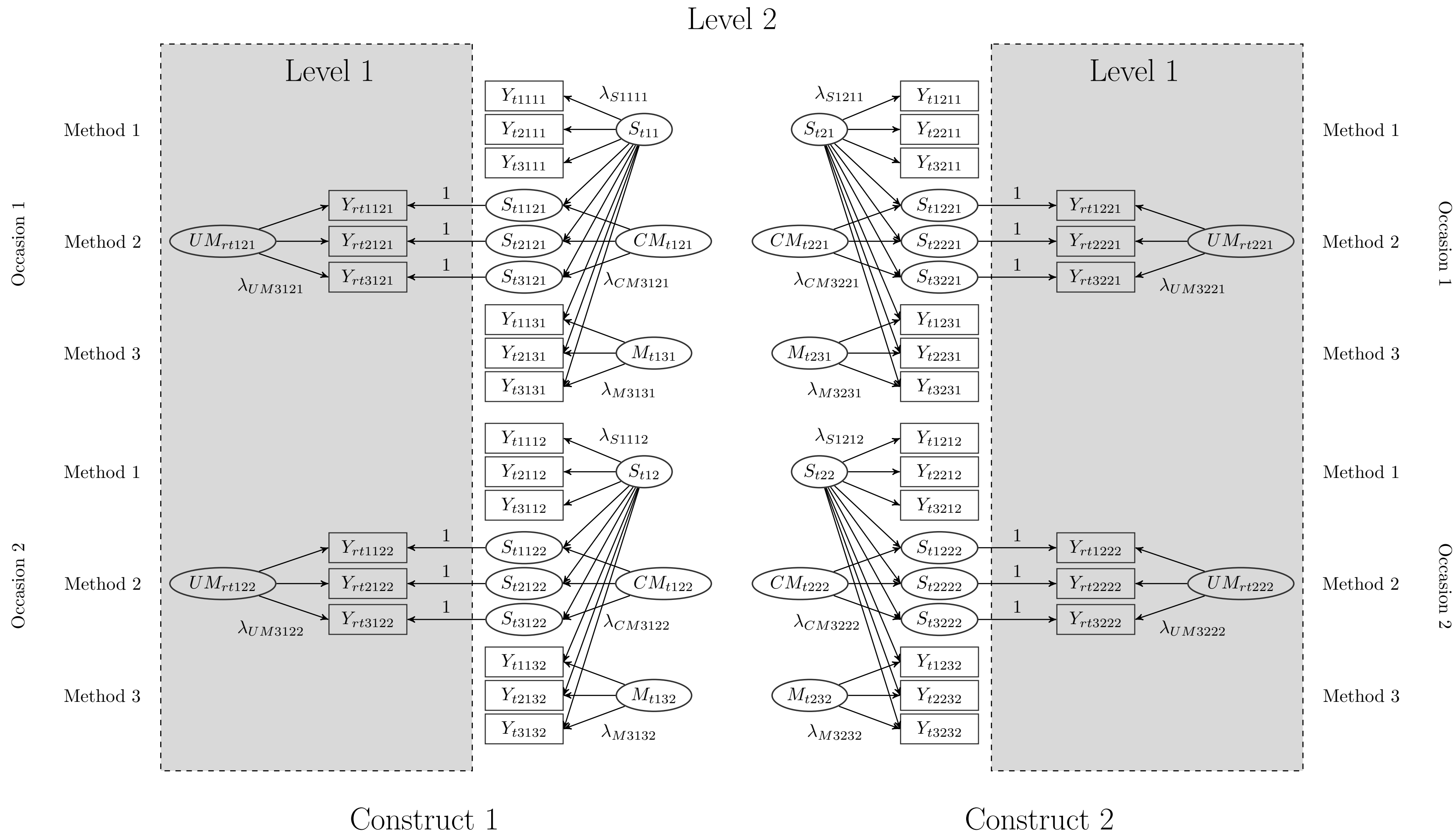

Figure 2.2: Path diagram of the Latent-State-Com graded response model with common latent state variables. The model is depicted for two structurally different methods and one set of interchangeable methods at two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For convenience, the constant indicator $k = 1$ has been dropped from the latent state variables ($S_{tjl} = S_{tj1l}$). For the sake of clarity, correlations between latent variables are omitted and loading parameters are only shown for exemplary indicators. Correlations that are not permissible by definition of the LS-Com GRM are correlations between the latent state variables and the latent (common) methods variables of the same construct j and occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method variable; M: method variable; S: latent state variable; UM: unique method variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l.

2.4 Formal Definition of the LS-Com GRM

In the following the LS-Com GRM is formally defined building on the definition of the graded response model (Samejima, 1969, 2010) and the LS-Com model for continuous indicators (Koch, 2013). The definition is based on stochastic measurement theory according to the approach by Eid (2000), Koch, Eid, and Lochner (in press), Steyer and Eid (2001) as well as Steyer (1988). The model is defined for two structurally different methods and one set of interchangeable methods. Adding additional methods is straightforward and does not need further definitions.

Definition 2.1. LS-Com GRM

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ *and* $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ *on a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *are variables of an LS-Com graded response model if the following conditions hold:*

(a) $(\Omega, \mathcal{A}, \mathcal{P})$ *is a probability space such that*

$$\Omega = \Omega_T \times \Omega_{TS_1} \times \ldots \times \Omega_{TS_l} \times \ldots \times \Omega_{TS_f} \times \Omega_R \times \Omega_{R_2 S_1} \times \ldots \times \Omega_{R_k S_l} \times \ldots \times \Omega_{R_e S_f} \times \Omega_O$$

where

$$\begin{aligned} \Omega_O &= \Omega_{O_{11}} \times \ldots \times \Omega_{O_{kl}} \times \ldots \times \Omega_{O_{ef}}\ , & \\ \Omega_{O_{kl}} &= A_{1kl} \times \ldots \times A_{jkl} \times \ldots \times A_{dkl}\ , & \forall\, k, l, \\ A_{jkl} &= O_{1jkl} \times \ldots \times O_{ijkl} \times \ldots \times O_{c_j jkl}\ , & \forall\, j, k, l \end{aligned}$$

with $i \in I_j = \{1, \ldots, c_j\}$, $j \in J = \{1, \ldots, d\}$, $k \in K = \{1, \ldots, e\}$, *and* $l \in L = \{1, \ldots, f\}$.

(b) The projections $p_T : \Omega \to \Omega_T$, $p_{TS_l} : \Omega \to \Omega_{TS_l}$, $p_R : \Omega \to \Omega_R$, *and* $p_{R_k S_l} : \Omega \to \Omega_{R_k S_l}$ *are random variables on* $(\Omega, \mathcal{A}, \mathcal{P})$.

(c) Without loss of generality, the first method $(k = 1)$ *is selected as reference method and defined as a self-report variable in the following. The second method* $(k = 2)$ *refers to the set of interchangeable methods which serve as non-reference methods. All other methods* $(k > 2)$ *refer to structurally different methods (that are not self-reports) which serve as non-reference methods. Then the variables*

$$\begin{aligned} Y_{tij1l} &: \quad \Omega_T \times \Omega_{TS_l} \times \Omega_O \to S_{ij} & \\ Y_{rtij2l} &: \quad \Omega_T \times \Omega_{TS_l} \times \Omega_R \times \Omega_{R_2 S_l} \times \Omega_O \to S_{ij} & \\ Y_{tijkl} &: \quad \Omega_T \times \Omega_{TS_l} \times \Omega_{R_k S_l} \times \Omega_O \to S_{ij} & k > 2 \end{aligned}$$

with i, j, k, l *as in (a), are random variables on* $(\Omega, \mathcal{A}, \mathcal{P})$ *that map the results of the random experiment onto the set* $S_{ij} \in \mathbb{N}_0$, $S_{ij} = \{0, \ldots, q_{ij} - 1\}$, *where* q_{ij} *is the number of response categories of item i of construct j.*

(d) Then the latent response variables π_{tsijkl} *and* $\pi_{rtsij2l}$ *are random variables on* $(\Omega, \mathcal{A}, \mathcal{P})$ *defined by:*

$$\pi_{tsij1l} := \Phi^{-1}[P(Y_{tij1l} \geq s \mid p_T, p_{TS_l})] \tag{2.4.1}$$
$$\pi_{rtsij2l} := \Phi^{-1}[P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2S_l})] \tag{2.4.2}$$
$$\pi_{tsijkl} := \Phi^{-1}[P(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_kS_l})] \qquad k > 2 \tag{2.4.3}$$

with Φ denoting the cumulative distribution function of the standard normal distribution, $s \in S_{ij}$ and

$$P(Y_{tij1l} \geq s \mid p_T, p_{TS_l}) = \mathbb{E}[I_{\{Y_{tij1l} \geq s\}} \mid p_T, p_{TS_l}] \tag{2.4.4}$$
$$P(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2S_l}) = \mathbb{E}[I_{\{Y_{rtij2l} \geq s\}} \mid p_T, p_{TS_l}, p_R, p_{R_2S_l}] \tag{2.4.5}$$
$$P(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_kS_l}) = \mathbb{E}[I_{\{Y_{tijkl} \geq s\}} \mid p_T, p_{TS_l}, p_{R_kS_l}], \qquad k > 2 \tag{2.4.6}$$

(e) *Essential π_{tijkl}- and π_{rtijkl}-equivalence. For each (s,i,j,k,l), $s \in S_{ij}$, $i \in I_j = \{1,\ldots,c_j\}$, $j \in J = \{1,\ldots,d\}$, $k \in K = \{1,\ldots,e\}$, and $l \in L = \{1,\ldots,f\}$ there is a constant $\kappa_{sijkl} \in \mathbb{R}$ and latent variables π_{tijkl} and π_{rtij2l} such that:*

$$\pi_{tijkl} := \pi_{tsijkl} + \kappa_{sijkl} \qquad k \neq 2 \tag{2.4.7}$$
$$\pi_{rtij2l} := \pi_{rtsij2l} + \kappa_{sij2l} \tag{2.4.8}$$

where π_{tijkl} and π_{rtij2l} are common latent response variables of all latent response variables π_{tsijkl} and $\pi_{rtsij2l}$, respectively, belonging to the same item i of construct j measured by method k at time l.

(f) *Then, the following variables are random variables on $(\Omega, \mathcal{A}, \mathcal{P})$ with finite first- and second-order moments:*

Rater-level (Level 1):

$$UM_{rtij2l} = \pi_{rtij2l} - \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \tag{2.4.9}$$

Target-level (Level 2):

$$S_{tij1l} = \pi_{tij1l} \tag{2.4.10}$$
$$S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \tag{2.4.11}$$
$$S_{tijkl} = \pi_{tijkl} \qquad \forall\ k > 2, \tag{2.4.12}$$
$$CM_{tij2l} = S_{tij2l} - \mathbb{E}[S_{tij2l} \mid S_{tij1l}] \tag{2.4.13}$$
$$M_{tijkl} = S_{tijkl} - \mathbb{E}[S_{tijkl} \mid S_{tij1l}] \qquad \forall\ k > 2, \tag{2.4.14}$$

(g) *For each construct j, measured by a non-reference method $(k \neq 1)$ on occasion of measurement l with item i, there are constants $\alpha_{ijkl} \in \mathbb{R}$ and $\lambda_{Sijkl} \in \mathbb{R}^+$ such that*

$$\mathbb{E}[S_{tijkl} \mid S_{tij1l}] = \alpha_{ijkl} + \lambda_{Sijkl} S_{tij1l}. \tag{2.4.15}$$

(h) *For each construct j, measured by an interchangeable non-reference method* $(k = 2)$ *on occasion of measurement l and for each pair* $(i,i') \in I_j \times I_j$, $(i \neq i')$, *there is a constant* $\lambda_{CMii'j2l} \in \mathbb{R}^+$ *such that*

$$CM_{tij2l} = \lambda_{CMii'j2l} CM_{ti'j2l}. \tag{2.4.16}$$

(i) *For each construct j, measured by an interchangeable non-reference method* $(k = 2)$ *on occasion of measurement l and for each pair* $(i,i') \in I_j \times I_j$, $(i \neq i')$, *there is a constant* $\lambda_{UMii'j2l} \in \mathbb{R}^+$ *such that*

$$UM_{rtij2l} = \lambda_{UMii'j2l} UM_{rti'j2l}. \tag{2.4.17}$$

(j) *For each construct j, measured by a non-reference method* $(k > 2)$ *on occasion of measurement l and for each pair* $(i,i') \in I_j \times I_j$, $(i \neq i')$, *there is a constant* $\lambda_{Mii'jkl} \in \mathbb{R}^+$ *such that*

$$M_{tijkl} = \lambda_{Mii'jkl} M_{ti'jkl} \qquad \forall \ k > 2. \tag{2.4.18}$$

Remarks. The indices used in the above definition stand for: r for rater, t for target, i for indicator, j for construct, k for method, and l for the occasion of measurement. The index k represents the type of method that is used (reference vs. non-reference method, self-report vs. informant report, structurally different vs. interchangeable method). The model is defined for one set of interchangeable methods only, denoted by $k = 2$. The index r indicates that a variable is measured on the within-level (level-1). Only the interchangeable methods are measured on the within-level, as they are the result of a multi-stage sampling procedure. That is, the interchangeable rater is sampled from a set of raters for a specific target. In contrast, the structurally different raters are fixed given the target. Hence, there is no additional sampling step for the structurally different raters once a target has been selected. However, also the responses of the structurally different raters are observed in rater-specific situations, and thus depend on the projection $p_{R_kS_l}$. Note that this model definition of the GRM deviate from their continuous indicator model counterparts in that the latter did not contain rater-specific situations for the structurally different raters and hence did not include the projection $p_{R_kS_l}$ for $k > 2$.
In order to distinguish between the rater-situations the subscript k is used. The set of rater-situations on measurement occasion l is assumed to include all the situations any of the raters could potentially encounter.
The variables π_{tij1l} can as well be represented as a composition of the mappings $(p_T, p_{TS_l}) : \Omega \rightarrow \Omega_T \times \Omega_{TS_l}$ and $\varphi_{tij1l} : \Omega_T \times \Omega_{TS_l} \rightarrow \mathbb{R}$, that is $\pi_{tij1l} : \varphi_{tij1l}(p_T, p_{TS_l})$ is a (p_T, p_{TS_l})-measurable function (Steyer & Nagel, 2017, pp. 57-58). Analogously, the variable π_{rtij2l} can be represented as the composite mapping of $(p_T, p_{TS_l}, p_R, p_{R_2S_l}) : \Omega \rightarrow \Omega_T \times \Omega_{TS_l} \times \Omega_R \times \Omega_{R_2S_l}$ and $\varphi_{rtij2l} : \Omega_T \times \Omega_{TS_l} \times \Omega_R \times \Omega_{R_2S_l} \rightarrow \mathbb{R}$, that is $\pi_{rtij2l} : \varphi_{rtij2l}(p_T, p_{TS_l}, p_R, p_{R_2S_l})$ is a $(p_T, p_{TS_l}, p_R, p_{R_2S_l})$- measurable function. In the same manner, for $k > 2$, $\pi_{tijkl} : \varphi_{tijkl}(p_T, p_{TS_l}, p_{R_kS_l})$ is a $(p_T, p_{TS_l}, p_{R_kS_l})$- measurable function, with $(p_T, p_{TS_l}, p_{R_kS_l}) : \Omega \rightarrow \Omega_T \times \Omega_{TS_l} \times \Omega_{R_kS_l}$ and $\varphi_{tijkl} : \Omega_T \times \Omega_{TS_l} \times \Omega_{R_kS_l} \rightarrow \mathbb{R}$.
Equations (2.4.16) - (2.4.18) define the assumptions that all latent method variables CM_{tij2l}, UM_{rtij2l}, and M_{tijkl} belonging to the same construct, method, and measurement occasion are similarity transformations of each other, respectively. This assumption implies that the variables are perfectly correlated and can therefore be represented by common method factors (Koch, 2013). The existence of these common method factors is stated in the following theorem (cf. Koch, 2013).

Note that assumptions (2.4.17) and (2.4.18) are necessary for identifiability reasons, as a model with indicator-specific factors UM_{rtij2l} or M_{tijkl} would not be identified. This is not the case for the CM_{tij2l} variables, hence it is not necessary to impose assumption (2.4.16) to identify the model. Note that the variables CM_{tij2l}, UM_{rtij2l}, and M_{tijkl} are defined as latent residual variables and therefore have expectations of zero by definition. Hence, no additive constants are included in Equations (2.4.16) - (2.4.18). Note that, in contrast, the latent state variables S_{tijkl} do not have zero expectations by definition, and the latent regression in Equation (2.4.15) does include an intercept parameter α_{ijkl}. The coefficients α_{ijkl} and all of the coefficients κ_{sijkl} for the same i, j, k, and l are, however, not separately identifiable (see Section 2.6 for theorems on the uniqueness of the latent state and latent response variables, as well as Section 2.13 on identifiability conditions).

Theorem 2.1. ***(Existence)***

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ *and* $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ *are* (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})*-congeneric variables of an LS-Com GRM if and only if conditions (a) to (j) of Definition 2.1 hold. Then, for each* $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$, *there are real-valued random variables* CM_{tij2l}, UM_{rtij2l}, *and* M_{tijkl} *on a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *and* (λ_{CMij2l}, λ_{UMij2l}, λ_{Mijkl}) $\in \mathbb{R}^+$ *such that:*

$$CM_{tij2l} = \lambda_{CMij2l} CM_{tj2l}, \tag{2.4.19}$$
$$UM_{rtij2l} = \lambda_{UMij2l} UM_{rtj2l}, \tag{2.4.20}$$
$$M_{tijkl} = \lambda_{Mijkl} M_{tjkl} \qquad \forall \;\; k > 2. \tag{2.4.21}$$

Remarks. The existence of the common factors CM_{tijkl}, UM_{rtijkl}, and M_{tijkl} follows directly from Equations (2.4.16) - (2.4.18) (i.e., Assumptions (h) - (j)) of Definition 2.1. Proofs of the existence of these latent variables were given by Koch (2013) and shall not be repeated here. The term *common* refers to the fact that each factor is assumed to be common to all indicators that belong to the same construct, the same method, and the same occasion of measurement.

2.5 Uniqueness, admissible transformations and meaningful statements

It is apparent that the latent method variables CM_{tijkl}, UM_{rtijkl}, and M_{tijkl} are not uniquely defined. If an LS-Com GRM is defined with (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric variables, these variables and their respective coefficients λ_{CMij2l}, λ_{UMij2l}, and λ_{Mijkl} are uniquely defined only up to similarity transformations, that is, up to the multiplication with a positive real number. A detailed theorem and proofs on admissible transformations and uniqueness of the common method variables are given by Koch (2013) for the LS-Com model with continuous indicator variables. These apply in the same manner to the LS-Com GRM. However, in the LS-Com GRM also the common latent response variables π_{rtijkl} and π_{tijkl} and their coefficients κ_{sijkl} as well as α_{ijkl} are not uniquely defined. A comprehensive theorem on the uniqueness of the latent variables is therefore given in the following theorem (cf., Eid, 1995; Koch, 2013).

Theorem 2.2. ***(Admissible transformations and uniqueness)***

1. Admissible Transformations
Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}),\ \boldsymbol{\pi}_{\mathbf{rt}},\ \boldsymbol{\pi}_{\mathbf{t}},\ \boldsymbol{\kappa},\ \boldsymbol{\lambda}_{\mathbf{S}},\ \boldsymbol{\alpha},\ \mathbf{UM_{rt}},\ \mathbf{CM_t},\ \mathbf{M_t},\ \boldsymbol{\lambda}_{\mathbf{UM}},\ \boldsymbol{\lambda}_{\mathbf{CM}},\ \boldsymbol{\lambda}_{\mathbf{M}}\rangle$ *be an LS-Com GRM with:*

$$\boldsymbol{\pi}_{\mathbf{rt}} = (\pi_{rt1121}, \ldots, \pi_{rtij2l}, \ldots, \pi_{rtc_d d2f})^T \tag{2.5.1}$$
$$\boldsymbol{\pi}_{\mathbf{t}} = (\pi_{t1111}, \ldots, \pi_{tijkl}, \ldots, \pi_{tc_d def})^T \qquad k \neq 2 \tag{2.5.2}$$
$$\boldsymbol{\kappa} = (\kappa_{11111}, \ldots, \kappa_{sijkl}, \ldots, \kappa_{(q_{c_d d}-1)c_d def})^T \tag{2.5.3}$$
$$\boldsymbol{\lambda}_{\mathbf{S}} = (\lambda_{S1111}, \ldots, \lambda_{Sijkl}, \ldots, \lambda_{Sc_d def})^T \tag{2.5.4}$$
$$\boldsymbol{\alpha} = (\alpha_{1121}, \ldots, \alpha_{ijkl}, \ldots, \alpha_{c_d def})^T \qquad k \neq 1 \tag{2.5.5}$$
$$\mathbf{UM_{rt}} = (UM_{rt121}, \ldots, UM_{rtj2l}, \ldots, UM_{rtd2f})^T \tag{2.5.6}$$
$$\mathbf{CM_t} = (CM_{t121}, \ldots, CM_{tj2l}, \ldots, CM_{td2f})^T \tag{2.5.7}$$
$$\mathbf{M_t} = (M_{t131}, \ldots, M_{tjkl}, \ldots, M_{tdef})^T \qquad k > 2 \tag{2.5.8}$$
$$\boldsymbol{\lambda}_{\mathbf{UM}} = (\lambda_{UM1121}, \ldots, \lambda_{UMij2l}, \ldots, \lambda_{UMc_d d2f})^T \tag{2.5.9}$$
$$\boldsymbol{\lambda}_{\mathbf{CM}} = (\lambda_{CM1121}, \ldots, \lambda_{CMij2l}, \ldots, \lambda_{CMc_d d2f})^T \tag{2.5.10}$$
$$\boldsymbol{\lambda}_{\mathbf{M}} = (\lambda_{M1131}, \ldots, \lambda_{Mijkl}, \ldots, \lambda_{Mc_d def})^T \qquad k > 2 \tag{2.5.11}$$

If for all $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$*:*

$$\pi'_{tij1l} = \pi_{tij1l} + \gamma_{ij1l} \tag{2.5.12}$$
$$\kappa'_{sijkl} = \kappa_{sijkl} + \gamma_{ijkl} \tag{2.5.13}$$
$$\alpha'_{ijkl} = \alpha_{ijkl} + \gamma_{ijkl} - \lambda_{Sijkl}\gamma_{ij1l} \qquad k \neq 1 \tag{2.5.14}$$
$$\lambda'_{UMij2l} = \lambda_{UMij2l}/\beta_{UMj2l} \tag{2.5.15}$$
$$\lambda'_{CMij2l} = \lambda_{CMij2l}/\beta_{CMj2l} \tag{2.5.16}$$
$$\lambda'_{Mijkl} = \lambda_{Mijkl}/\beta_{Mjkl} \tag{2.5.17}$$
$$UM'_{rtj2l} = \beta_{UMj2l} UM_{rtj2l} \tag{2.5.18}$$
$$CM'_{tj2l} = \beta_{CMj2l} CM_{tj2l} \tag{2.5.19}$$
$$M'_{tjkl} = \beta_{Mjkl} CM_{tjkl} \tag{2.5.20}$$

where β_{UMj2l}, β_{CMj2l}, β_{Mjkl}, $\gamma_{ijkl} \in \mathbb{R}$, *and* β_{UMj2l}, β_{CMj2l} *and* $\beta_{Mjkl} > 0$, *and*

$$\pi'_{rtij2l} = \alpha'_{ij2l} + \lambda_{Sij2l}\pi'_{tij1l} + \lambda'_{CMij2l} CM'_{tj2l} + \lambda'_{UMij2l} UM'_{rtj2l} \tag{2.5.21}$$
$$\pi'_{tijkl} = \alpha'_{ijkl} + \lambda_{Sijkl}\pi'_{tij1l} + \lambda'_{Mijkl} M'_{tjkl} \qquad k > 2 \tag{2.5.22}$$

Then $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P}),\ \boldsymbol{\pi}'_{\mathbf{rt}},\ \boldsymbol{\pi}'_{\mathbf{t}},\ \boldsymbol{\kappa}',\ \boldsymbol{\lambda}_{\mathbf{S}},\ \boldsymbol{\alpha}',\ \mathbf{UM}'_{\mathbf{rt}},\ \mathbf{CM}'_{\mathbf{t}},\ \mathbf{M}'_{\mathbf{t}},\ \boldsymbol{\lambda}'_{\mathbf{UM}},\ \boldsymbol{\lambda}'_{\mathbf{CM}},\ \boldsymbol{\lambda}'_{\mathbf{M}}\rangle$ *is an LS-Com GRM, too, with*

$$\boldsymbol{\pi}'_{\mathbf{rt}} = (\pi'_{rt1121}, \ldots, \pi'_{rtij2l}, \ldots, \pi'_{rtc_d2f})^T \tag{2.5.23}$$
$$\boldsymbol{\pi}'_{\mathbf{t}} = (\pi'_{t1111}, \ldots, \pi'_{tijkl}, \ldots, \pi'_{tc_ddef})^T \quad k \neq 2 \tag{2.5.24}$$
$$\boldsymbol{\kappa}' = (\kappa'_{11111}, \ldots, \kappa'_{sijkl}, \ldots, \kappa'_{(q_{cd}-1)c_ddef})^T \tag{2.5.25}$$
$$\boldsymbol{\lambda}_{\mathbf{S}} = (\lambda_{S1111}, \ldots, \lambda_{Sijkl}, \ldots, \lambda_{Sc_ddef})^T \tag{2.5.26}$$
$$\boldsymbol{\alpha}' = (\alpha'_{1121}, \ldots, \alpha'_{ijkl}, \ldots, \alpha'_{c_ddef})^T \quad k \neq 1 \tag{2.5.27}$$

$$\mathbf{UM}'_{\mathbf{rt}} = (UM'_{rt121}, \ldots, UM'_{rtj2l}, \ldots, UM'_{rtd2f})^T \tag{2.5.28}$$
$$\mathbf{CM}'_{\mathbf{t}} = (CM'_{t121}, \ldots, CM'_{tj2l}, \ldots, CM'_{td2f})^T \tag{2.5.29}$$
$$\mathbf{M}'_{\mathbf{t}} = (M'_{t131}, \ldots, M'_{tjkl}, \ldots, M'_{tdef})^T \quad k > 2 \tag{2.5.30}$$

$$\boldsymbol{\lambda}'_{\mathbf{UM}} = (\lambda'_{UM1121}, \ldots, \lambda'_{UMij2l}, \ldots, \lambda'_{UMc_dd2f})^T \tag{2.5.31}$$
$$\boldsymbol{\lambda}'_{\mathbf{CM}} = (\lambda'_{CM1121}, \ldots, \lambda'_{CMij2l}, \ldots, \lambda'_{CMc_dd2f})^T \tag{2.5.32}$$
$$\boldsymbol{\lambda}'_{\mathbf{M}} = (\lambda'_{M1131}, \ldots, \lambda'_{Mijkl}, \ldots, \lambda'_{Mc_ddef})^T \quad k > 2 \tag{2.5.33}$$

2. Uniqueness
If both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{\mathbf{rt}}$, $\boldsymbol{\pi}_{\mathbf{t}}$, $\boldsymbol{\kappa}$, $\boldsymbol{\lambda}_{\mathbf{S}}$, $\boldsymbol{\alpha}$, $\mathbf{UM}_{\mathbf{rt}}$, $\mathbf{CM}_{\mathbf{t}}$, $\mathbf{M}_{\mathbf{t}}$, $\boldsymbol{\lambda}_{\mathbf{UM}}$, $\boldsymbol{\lambda}_{\mathbf{CM}}$, $\boldsymbol{\lambda}_{\mathbf{M}}\rangle$ *and* $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{\mathbf{rt}}$, $\boldsymbol{\pi}'_{\mathbf{t}}$, $\boldsymbol{\kappa}'$, $\boldsymbol{\lambda}_{\mathbf{S}}$, $\boldsymbol{\alpha}'$, $\mathbf{UM}'_{\mathbf{rt}}$, $\mathbf{CM}'_{\mathbf{t}}$, $\mathbf{M}'_{\mathbf{t}}$, $\boldsymbol{\lambda}'_{\mathbf{UM}}$, $\boldsymbol{\lambda}'_{\mathbf{CM}}$, $\boldsymbol{\lambda}'_{\mathbf{M}}\rangle$ *are LS-Com GRMs, then for each* $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$ *there are* $\gamma_{ijkl} \in \mathbb{R}$, *and* β_{UMj2l}, β_{CMj2l}, $\beta_{Mjkl} \in \mathbb{R}^+$, *such that Equations* (2.5.12) *to* (2.5.22) *hold.*

Remarks. Equations (2.5.12) - (2.5.22) define the transformations of the latent variables and their parameters that would yield an equivalent LS-Com GRM. Note that Equations (2.5.21) and (2.5.22) stating the transformations for the latent response variables π'_{rtij2l} and π'_{tijkl}, $k > 2$, are equivalent to the measurement equations of these variables in the transformed model $\mathcal{M}'$. They denote that these variables change by a positive real constant γ_{ijkl} due to the changes in the latent response variable π_{tij1l} and the intercepts α_{ijkl} given in Equations (2.5.12) and (2.5.14).
As the common method factors and their corresponding loading parameters are uniquely defined only up to similarity transformations, admissible transformations of these factors and loadings are the multiplications with positive real numbers. These admissible transformations determine which meaningful statements (statements that remain invariant under admissible transformations) can be made regarding the latent variables and their coefficients in the LS-Com GRM. As the multiplication with positive real numbers is an admissible transformation of the latent method factors CM_{tijkl}, UM_{rtijkl}, and M_{tijkl} and their corresponding factor loadings, statements regarding the absolute value of the parameters are not meaningful. Meaningful statements are statements regarding the ratio of specific values of the factor loadings or the ratio of the values of latent method factors (see Geiser, 2008; Koch, 2013). That is, they are measured on a ratio scale. Possible meaningful statements on method factors as well as on their factor loadings shall be illustrated with the example of the unique method loadings. Let both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{\mathbf{rt}}$, $\boldsymbol{\pi}_{\mathbf{t}}$, $\boldsymbol{\kappa}$, $\boldsymbol{\lambda}_{\mathbf{S}}$, $\boldsymbol{\alpha}$, $\mathbf{UM}_{\mathbf{rt}}$, $\mathbf{CM}_{\mathbf{t}}$, $\mathbf{M}_{\mathbf{t}}$, $\boldsymbol{\lambda}_{\mathbf{UM}}$, $\boldsymbol{\lambda}_{\mathbf{CM}}$, $\boldsymbol{\lambda}_{\mathbf{M}}\rangle$ and $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{\mathbf{rt}}$, $\boldsymbol{\pi}'_{\mathbf{t}}$, $\boldsymbol{\kappa}'$, $\boldsymbol{\lambda}_{\mathbf{S}}$, $\boldsymbol{\alpha}'$, $\mathbf{UM}'_{\mathbf{rt}}$, $\mathbf{CM}'_{\mathbf{t}}$, $\mathbf{M}'_{\mathbf{t}}$, $\boldsymbol{\lambda}'_{\mathbf{UM}}$, $\boldsymbol{\lambda}'_{\mathbf{CM}}$, $\boldsymbol{\lambda}'_{\mathbf{M}}\rangle$ be LS-Com GRMs defined by Equations (2.5.1) to (2.5.33). Then, for $\omega_1, \omega_2 \in \Omega$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$:

$$\frac{\lambda_{UMij2l}}{\lambda_{UMi'j2l}} = \frac{\lambda'_{UMij2l}}{\lambda'_{UMi'j2l}} \tag{2.5.34}$$

and

$$\frac{UM_{rtj2l}(\omega_1)}{UM_{rtj2l}(\omega_2)} = \frac{UM'_{rtj2l}(\omega_1)}{UM'_{rtj2l}(\omega_2)} \tag{2.5.35}$$

Thus, statements regarding the ratio of specific values of the factor loadings or the ratio of values of latent method factors are meaningful. Values on the method factors of different targets can therefore be compared using their ratio. Meaningful statements with regard to CM_{rtij2l}, M_{tijkl}, λ_{CMij2l}, and λ_{Mijkl} can be made in the same manner. Furthermore, the products $\lambda^2_{CMij2l}Var(CM_{tj2l})$, $\lambda^2_{UMij2l}Var(UM_{rtj2l})$, and $\lambda^2_{Mijkl}Var(M_{tjkl})$ are invariant under similarity transformations, as, e.g.,

$$\begin{aligned}\lambda^2_{UMij2l}Var(UM_{rtj2l}) &= \frac{\lambda^2_{UMij2l}}{\beta^2_{UMj2l}}\beta^2_{UMj2l}Var(UM_{rtj2l}) \\ &= \lambda'^2_{UMij2l}Var(UM'_{rtj2l})\end{aligned} \tag{2.5.36}$$

Hence, any statement with respect to the ratio of variances are meaningful. This property ensures the meaningfulness of statements concerning variance components such as consistency and method specificity coefficients (see Section 2.11). Also, statements concerning latent correlations between method factors are meaningful, as, for $j, j' \in J$, and $l, l' \in L$ (Steyer & Nagel, 2017, remark 7.21, p. 243):

$$Corr(UM_{j2l}, UM_{j'2l'}) = Corr(UM'_{j2l}, UM'_{j'2l'}). \tag{2.5.37}$$

Proofs on the uniqueness, admissible transformations and meaningful statements concerning the latent method factors and their coefficients can be found in Koch (2013).
From Equations (2.5.12) and (2.5.13) it follows that the latent common response variables π_{tij1l} and their respective threshold parameters κ_{sij1l} are uniquely defined only up to translations. That is, they are measured on a difference scale. The same holds for the latent response variables π_{tijkl}, $k > 2$, and π_{rtij2l} and their threshold parameters κ_{sijkl} by Equations (2.5.21) - (2.5.22) and (2.5.13).

Proof. *Admissible transformations and uniqueness of the latent common response variables.*

Let π'_{tijkl}, π'_{rtij2l}, and κ'_{sijkl} be defined as given by Equations (2.5.12), (2.5.13), (2.5.21), *and* (2.5.22). *Then, it holds that:*

$$\begin{aligned}\pi_{tsijkl} &= \pi_{tijkl} - \kappa_{sijkl} = (\pi_{tijkl} + \gamma_{ijkl}) - (\kappa_{sijkl} + \gamma_{ijkl}) \\ &= \pi'_{tijkl} - \kappa'_{sijkl} \qquad \forall\, k \neq 2\end{aligned}$$

and

$$\begin{aligned}\pi_{rtsij2l} &= \pi_{rtij2l} - \kappa_{sij2l} = (\pi_{rtij2l} + \gamma_{ij2l}) - (\kappa_{sij2l} + \gamma_{ij2l}) \\ &= \pi'_{rtij2l} - \kappa'_{sij2l}.\end{aligned}$$

Let both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi_{rt}}$, $\boldsymbol{\pi_t}$, $\boldsymbol{\kappa}$, $\boldsymbol{\lambda_S}$, $\boldsymbol{\alpha}$, $\mathbf{UM_{rt}}$, $\mathbf{CM_t}$, $\mathbf{M_t}$, $\boldsymbol{\lambda_{UM}}$, $\boldsymbol{\lambda_{CM}}$, $\boldsymbol{\lambda_M}\rangle$ and $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi'_{rt}}$, $\boldsymbol{\pi'_t}$, $\boldsymbol{\kappa'}$, $\boldsymbol{\lambda_S}$, $\boldsymbol{\alpha'}$, $\mathbf{UM'_{rt}}$, $\mathbf{CM'_t}$, $\mathbf{M'_t}$, $\boldsymbol{\lambda'_{UM}}$, $\boldsymbol{\lambda'_{CM}}$, $\boldsymbol{\lambda'_M}\rangle$ be LS-Com GRMs. Then it has to hold that $\pi_{tijkl} - \kappa_{sijkl} = \pi'_{tijkl} - \kappa'_{sijkl}$ and $\pi_{rtij2l} - \kappa_{sij2l} = \pi'_{rtij2l} - \kappa'_{sij2l}$ for all $s \in S_{ij}$, $i \in I_j$, $j \in J$, $k \in K, k \neq 2$, and $l \in L$. It follows that $\pi'_{tijkl} = \pi_{tijkl} - \kappa_{sijkl} + \kappa'_{sijkl}$, $k \neq 2$, and $\pi'_{rtij2l} = \pi_{rtij2l} - \kappa_{sij2l} + \kappa'_{sij2l}$. As the difference $\kappa'_{sijkl} - \kappa_{sijkl}$ has to be the same over all $s \in S_{ij}$ for each $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, one can define γ_{ijkl} as $\gamma_{ijkl} = \kappa'_{sijkl} - \kappa_{sijkl}$.

□

Therefore, meaningful statements regarding the common latent response variables π_{tijkl} and π_{rtij2l} are statements on their differences: for ω_1, $\omega_2 \in \Omega$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, it holds that

$$\pi_{tijkl}(\omega_1) - \pi_{tijkl}(\omega_2) = \pi'_{tijkl}(\omega_1) - \pi'_{tijkl}(\omega_2) \tag{2.5.38}$$

and

$$\pi_{rtij2l}(\omega_1) - \pi_{rtij2l}(\omega_2) = \pi'_{rtij2l}(\omega_1) - \pi'_{rtij2l}(\omega_2) \tag{2.5.39}$$

as, e.g., for π_{tijkl}:

$$\begin{aligned} \pi'_{tijkl}(\omega_1) - \pi'_{tijkl}(\omega_2) &= \big(\pi_{tijkl}(\omega_1) + \gamma_{ijkl}\big) - \big(\pi_{tijkl}(\omega_2) + \gamma_{ijkl}\big) \\ &= \pi_{tijkl}(\omega_1) - \pi_{tijkl}(\omega_2) \end{aligned} \tag{2.5.40}$$

From Equations (2.5.38) and (2.5.39) it follows that statements about the change in the latent response variables π_{tijkl} and π_{rtijkl} between different occasions are only meaningful for the differences between persons, that is, for $\omega_1, \omega_2 \in \Omega$, $i \in I_j$, $j \in J$, $k \in K$, and $l, l' \in L$, it holds that

$$\begin{aligned} &\big[\pi_{tijkl}(\omega_1) - \pi_{tijkl'}(\omega_1)\big] - \big[\pi_{tijkl}(\omega_2) - \pi_{tijkl'}(\omega_2)\big] \\ =&\big[\pi'_{tijkl}(\omega_1) - \pi'_{tijkl'}(\omega_1)\big] - \big[\pi'_{tijkl}(\omega_2) - \pi'_{tijkl'}(\omega_2)\big] \end{aligned} \tag{2.5.41}$$

Note that every result on the uniqueness, admissible transformations and meaningful statements regarding the latent response variables π_{tijkl} and π_{rtijkl} also apply to the latent state variables $S_{tij1l} = \pi_{tij1l}$, $S_{tijkl} = \pi_{tijkl} \ \ \forall\, k > 2$, and $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$, by definition. Thus, meaningful statements about the change of a person's value on the latent state variable S_{tijkl} between two measurement occasions l and l' can be made only compared to another person's change on these latent state variables.
Statements concerning latent covariances and correlations between the latent state factors S_{tij1l} are meaningful, as the addition of constants does not influence the covariance structure. That is, for all $i, i' \in I_j$, $j, j' \in J$, and $l, l' \in L$, it holds that

$$Corr(S'_{tij1l}, S'_{ti'j'1l'}) = Corr(S_{tij1l} + \gamma_{ij1l}, S_{ti'j'1l'} + \gamma_{i'j'1l'}) = Corr(S_{tij1l}, S_{ti'j'1l'}) \tag{2.5.42}$$

as γ_{ij1l} and $\gamma_{i'j'1l'}$ are constants. For the threshold parameters κ_{sijkl}, meaningful statements refer to differences between the thresholds of one item i, i.e., for all $s, s' \in S_{ij}$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, it holds that:

$$\kappa_{sijkl} - \kappa_{s'ijkl} = \kappa'_{sijkl} - \kappa'_{s'ijkl} \tag{2.5.43}$$

and for $i, i' \in I_j$, $j, j' \in J$, $k, k' \in K$, and $l, l' \in L$ with $(i, j, k, l) \neq (i, j, k, l)'$

$$(\kappa_{sijkl} - \kappa_{s'ijkl}) - (\kappa_{s(ijkl)'} - \kappa_{s'(ijkl)'}) = (\kappa'_{sijkl} - \kappa'_{s'ijkl}) - (\kappa'_{s(ijkl)'} - \kappa'_{s'(ijkl)'}) \tag{2.5.44}$$

2.6 Common latent state factors and uniqueness of the latent response variables

In an analogous manner to the common method factors, common latent state factors S_{tj1l} can be construed. That is, the LS-Com GRM can be defined with common latent state factors S_{tj1l} or with indicator-specific latent state factors S_{tij1l}. The specification of common latent state factors is based on the assumption that the indicator-specific latent state factors S_{tij1l} of the reference method, pertaining to the same construct j and same occasion of measurement l, are perfectly correlated (S_{tij1l}-congenerity) and can therefore be expressed as:

$$S_{tij1l} = \lambda_{Sij1l}\big(S_{tj1l} + \delta_{ij1l}\big). \tag{2.6.1}$$

with $\delta_{ij1l} \in \mathbb{R}$ and $\lambda_{Sij1l} \in \mathbb{R}^+$, $\lambda_{Sij1l} > 0$. Note that the expectations of the latent state variables S_{tij1l} and S_{tj1l} are, in contrast to those of the common method factors, not zero by definition and the variables S_{tij1l} are hence positive linear functions of each other. It can be seen that the common latent state variables S_{tj1l} as well as the coefficients δ_{ij1l} are only uniquely defined up to linear transformations, while the coefficients λ_{Sij1l} are uniquely defined up to similarity transformations. That is, transforming the common latent state variable S_{tj1l} by $S'_{tj1l} = \beta_{Sj1l}S_{tj1l} + \gamma_{Sj1l}$ and the coefficients λ_{Sij1l} and δ_{ij1l} by $\lambda'_{Sij1l} = \lambda_{Sij1l}/\beta_{Sj1l}$ and $\delta'_{ij1l} = \beta_{Sj1l}\delta_{ij1l} - \gamma_{Sj1l}$ yields

$$\begin{aligned} S_{tij1l} &= \lambda_{Sij1l}\big(S_{tj1l} + \delta_{ij1l}\big) \\ &= \frac{\lambda_{Sij1l}}{\beta_{Sj1l}}\big[\big(\beta_{Sj1l}S_{tj1l} + \gamma_{Sj1l}\big) + \big(\beta_{Sj1l}\delta_{ij1l} - \gamma_{Sj1l}\big)\big] \\ &= \lambda'_{Sij1l}\big(S'_{tj1l} + \delta'_{ij1l}\big) \end{aligned} \tag{2.6.2}$$

Meaningful statements regarding the common latent state factors are therefore statements about the ratio of differences between different values of S_{tij1l}, that is, for ω_1, ω_2, ω_3, $\omega_4 \in \Omega$, $t \in T$, $j \in J$ and $l \in L$, it holds that

$$\frac{S_{tj1l}(\omega_1) - S_{tj1l}(\omega_2)}{S_{tj1l}(\omega_3) - S_{tj1l}(\omega_4)} = \frac{S'_{tj1l}(\omega_1) - S'_{tj1l}(\omega_2)}{S'_{tj1l}(\omega_3) - S'_{tj1l}(\omega_4)} \tag{2.6.3}$$

However, as $S_{tij1l} = \pi_{tij1l}$ and κ_{sij1l} are only uniquely defined up to translations, the coefficients δ_{ij1l} and all of the coefficients κ_{sij1l} for the same i, j, and l are not separately identifiable. For further restrictions imposed on the mean structure, as well as on the coefficients α_{ijkl} and κ_{sijkl} due to identifiability considerations refer to Sections 2.12 and 2.13.

2.7 True score variables

As introduced by Eid (1995), latent true score variables can be defined for ordered categorical variables in the context of graded response models. These true score variables are defined as the expected value of the response given the target, the rater and their respective situations. It represents a continuous latent variable that is a function of the latent response variable π and the item difficulties κ. In analogy to the dichotomous response case, the curve describing the dependency of the latent true score variable on the latent response variable π has been termed item characteristic (Eid, 1995). The latent true score variables are monotonically increasing non-linear functions of the latent response variables. They are defined in the following Definition 2.2.

Definition 2.2. (True Score variables)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\pi_{\mathbf{rt}}$, $\pi_{\mathbf{t}}$, κ, α, $\mathbf{UM_{rt}}$, $\mathbf{CM_t}$, $\mathbf{M_t}$, $\lambda_{\mathbf{S}}$, $\lambda_{\mathbf{UM}}$, $\lambda_{\mathbf{CM}}$, $\lambda_{\mathbf{M}}\rangle$ *be an LS-Com GRM of* (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})*-congeneric variables. Then, the latent variables*

$$\tau_{tij1l} = \mathbb{E}(Y_{tij1l} \mid p_T, p_{TS_l}) \tag{2.7.1}$$
$$\tau_{rtij2l} = \mathbb{E}(Y_{rtij2l} \mid p_T, p_{TS_l}, p_R, p_{R_2S_l}) \tag{2.7.2}$$
$$\tau_{tijkl} = \mathbb{E}(Y_{tijkl} \mid p_T, p_{TS_l}, p_{R_kS_l}) \qquad k > 2 \tag{2.7.3}$$

are called latent true score variables and are given by

$$\tau_{tij1l} = \sum_{s=1}^{q_{ij}-1} P\big(Y_{tij1l} \geq s \mid p_T, p_{TS_l}\big) = \sum_{s=1}^{q_{ij}-1} \Phi\big(\pi_{tij1l} - \kappa_{sij1l}\big) \tag{2.7.4}$$
$$\tau_{rtij2l} = \sum_{s=1}^{q_{ij}-1} P\big(Y_{rtij2l} \geq s \mid p_T, p_{TS_l}, p_R, p_{R_2S_l}\big) = \sum_{s=1}^{q_{ij}-1} \Phi\big(\pi_{rtij2l} - \kappa_{sij2l}\big) \tag{2.7.5}$$
$$\tau_{tijkl} = \sum_{s=1}^{q_{ij}-1} P\big(Y_{tijkl} \geq s \mid p_T, p_{TS_l}, p_{R_kS_l}\big) = \sum_{s=1}^{q_{ij}-1} \Phi\big(\pi_{tijkl} - \kappa_{sijkl}\big) \qquad k > 2. \tag{2.7.6}$$

Remarks. According to Equations (2.7.4) - (2.7.6), the true score variables are additively composed of the conditional probabilities that the observed variables Y_{tijkl} or Y_{rtij2l} take on a value $\geq s$. Equations (2.7.4) - (2.7.6) follow directly from the definition of the true score variables as conditional expectations of the observed variables in Equations (2.7.1) - (2.7.3) and general rules for random variables in $\mathbb{N}$ (Bauer, 2002, theorem 3.10, p. 17). A detailed proof analogous to the proof of Equations (2.7.4) - (2.7.6) was given by Eid (1995, pp. 207-208) for a comparable model.

2.8 Factor analytical representation

The LS-Com GRM presented above can also be represented as a factor model for ordinal data. This approach is based on the assumption of the existence of unobservable continuous variables Y^*_{tijkl}, $k \neq 2$, and Y^*_{rtij2l} underlying the observed ordered categorical variables Y_{tijkl} and Y_{rtij2l}, respectively. The variables Y^*_{tijkl} and Y^*_{rtij2l} cannot be defined on the probability space described above or expressed as functions of measurement-theoretically well-defined measures. However, given certain assumptions / conditions, the approaches are formally equivalent as they imply the same multivariate distribution of the observed variables Y. A formal proof of the equivalence of the two approaches was given by, e.g., Takane and De Leeuw (1987).
For observed variables taking on one of q_{ij} different values out of the set of possible categories $S_{ij} = \{0, \cdots, q_{ij} - 1\}$, the relation between Y^*_{tijkl} and Y_{tijkl}, $k \neq 2$, as well as Y^*_{rtij2l} and Y_{rtij2l} is given by the following measurement structure:

$$Y_{tijkl} = \begin{cases} 0 & \text{for } Y^*_{tijkl} \leq \kappa^*_{1ijkl} \\ s & \text{for } \kappa^*_{sijkl} < Y^*_{tijkl} \leq \kappa^*_{(s+1)ijkl} \quad \textit{with } 0 < s < q_{ij} - 1 \\ q_{ij} - 1 & \text{for } \kappa^*_{(q_{ij}-1)ijkl} < Y^*_{tijkl} \end{cases} \tag{2.8.1}$$

$$Y_{rtij2l} = \begin{cases} 0 & \text{for } Y^*_{rtij2l} \leq \kappa^*_{1ij2l} \\ s & \text{for } \kappa^*_{sij2l} < Y^*_{rtij2l} \leq \kappa^*_{(s+1)ij2l} \quad \textit{with } 0 < s < q_{ij} - 1 \\ q_{ij} - 1 & \text{for } \kappa^*_{(q_{ij}-1)ij2l} < Y^*_{rtij2l} \end{cases} \tag{2.8.2}$$

According to this measurement model, the variables Y_{tijkl} and Y_{rtij2l} result from a categorization of the variables Y^*_{tijkl} and Y^*_{rtij2l} that is determined by the threshold parameters κ^*_{sijkl}.
Under the condition that $\kappa^*_{sijkl} = \kappa_{sijkl}$, in the factor analytical representation of the LS-Com GRM,

the continuous variables Y^*_{tijkl} and Y^*_{rtij2l} are functions of the variables π_{tijkl} and π_{rtij2l}, respectively, and an error term (ε_{tijkl} or ε_{rtij2l}):

$$Y^*_{tijkl} = \pi_{tijkl} + \varepsilon_{tijkl} \tag{2.8.3}$$

$$Y^*_{rtij2l} = \pi_{rtij2l} + \varepsilon_{rtij2l} \tag{2.8.4}$$

with $\mathbb{E}(\varepsilon_{(r)tijkl}) = 0$. The expectation and variance of Y^*_{ijkl} are

$$\mu^*_{ijkl} = \mathbb{E}(\pi_{ijkl}) \tag{2.8.5}$$

$$\sigma^*_{ijkl} = \Psi_{ijkl} + \theta_{ijkl} \tag{2.8.6}$$

where Ψ_{ijkl} is the variance of π_{ijkl}, and θ_{ijkl} is the variance of $\varepsilon_{(r)tijkl}$.
The error terms ε_{tijkl} and ε_{rtij2l} are assumed to be conditionally multivariate normally distributed given π_{tijkl} and π_{rtij2l}, as well as unconditionally multivariate normally distributed. Assuming multivariate normally distributed latent factors and thereby response variables π_{tijkl} and π_{rtij2l}, this results in multivariate normally distributed variables Y^*_{tijkl} and Y^*_{rtij2l}. The variables Y^*_{tijkl} and Y^*_{rtij2l} are, as a consequence, also conditionally multivariate normally distributed given π_{tijkl} and π_{rtij2l}, respectively, with a mean of μ^*_{ijkl} and μ^*_{ij2l}.
As the variables Y^*_{tijkl} and Y^*_{rtij2l} are latent variables, their metric is not determined, giving rise to different possibilities of standardization (see, e.g., B. Muthén & Asparouhov, 2002). One common standardization for the expectation is to set $\mu^*_{ijkl} = 0$. Then, one possibility is to set the σ^*_{ijkl} to an arbitrary value, e.g., $\sigma^*_{ijkl} = 1$. Under this parameterization the residual variances θ are not free parameters to be estimated in the model, but are given by

$$\theta_{ijkl} = 1 - \Psi_{ijkl} \tag{2.8.7}$$

Another possibility of standardization is to fix the residual variances, e.g., $\theta_{ijkl} = 1$, resulting in

$$\sigma^*_{ijkl} = \Psi_{ijkl} + 1 \tag{2.8.8}$$

Using the assumption of conditionally normally distributed error terms ε_{tijkl} or ε_{rtij2l} given π_{tijkl} and π_{rtij2l}, the conditional distribution of the variables Y^*_{tijkl}, $k \neq 2$, can be expressed as

$$\begin{aligned} P\Big(Y^*_{tijkl} \geq \kappa^*_{sijkl}\Big|\pi_{tijkl}\Big) &= P\Big(\pi_{tijkl} + \varepsilon_{tijkl} \geq \kappa^*_{sijkl}\Big|\pi_{tijkl}\Big) \\ &= P\Big(\varepsilon_{tijkl}\ \theta_{ijkl}^{-1/2} \geq (\kappa^*_{sijkl} - \pi_{tijkl})\ \theta_{ijkl}^{-1/2}\Big|\pi_{tijkl}\Big) \\ &= \Phi\Big((\pi_{tijkl} - \kappa^*_{sijkl})\ \theta_{ijkl}^{-1/2}\Big) \end{aligned} \tag{2.8.9}$$

and analogously for Y^*_{rtij2l}

$$P\Big(Y^*_{rtij2l} \geq \kappa^*_{sij2l}\Big|\pi_{rtij2l}\Big) = \Phi\Big((\pi_{rtij2l} - \kappa^*_{sij2l})\ \theta_{ij2l}^{-1/2}\Big), \tag{2.8.10}$$

where θ_{ijkl} is the variance of ε_{ijkl}. Recall from Equations (2.2.10) and (2.2.11) that the probability functions in the LS-Com GRM are given by

$$P(Y_{rtij2l} \geq s \mid \pi_{rtij2l}) = \Phi(\pi_{rtij2l} - \kappa_{sij2l})$$

and, for $k \neq 2$,

$$P(Y_{tijkl} \geq s \mid \pi_{tijkl}) = \Phi(\pi_{tijkl} - \kappa_{sijkl}).$$

Hence, the parameters of the factor analytical approach can be transformed to yield the parameters of a GRM in the following way (Eid, 1995; B. Muthén & Asparouhov, 2002; Takane & De Leeuw, 1987):

$$\kappa_{sijkl} = \frac{\kappa^*_{sijkl}}{\sqrt{\theta_{ijkl}}} \tag{2.8.11}$$

and, analogously, denoting the loading parameters in the factor analytical representation by λ^*_{ijkl}, that is, e.g., for $k > 2$, $\pi_{tijkl} = \lambda^*_{S_{ij1l}} S_{tij1l} + \lambda^*_{M_{ijkl}} M_{tijkl}$

$$\lambda_{ijkl} = \frac{\lambda^*_{ijkl}}{\sqrt{\theta_{ijkl}}} \tag{2.8.12}$$

for all types of loading parameters λ_S, λ_{UM}, λ_{CM} and λ_M. It follows that the parameter estimates in the two approaches are equivalent, resulting in the same values for the loading parameters as well as thresholds $\kappa_{sijkl} = \kappa^*_{sijkl}$, if and only if $\theta_{ijkl} = 1$ for all $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$ (Eid, 1995). The relation between the two approaches depicted above allows for the estimation of the parameters of the GRM with programs for SEMs (Eid, 1995). Furthermore, identification rules for SEMs can be applied to the LS-Com GRM (Eid, 1996), with some model-specific modifications. Identification conditions can be derived from the univariate and bivariate marginal probability expressions in the factor analytical representation. In the bivariate case, assuming uni- and bivariate normality of the variables Y^*_{tijkl}, these are given by (B. Muthén & Asparouhov, 2002):

$$\begin{aligned} P\left(Y_{tijkl} \geq s\right) = P\left(Y^*_{tijkl} \geq \kappa^*_{sijkl}\right) &= 1 - \Phi\left(\frac{\kappa^*_{sijkl} - \mu^*_{ijkl}}{\sigma^*_{ijkl}}\right) = \Phi\left(\frac{\mu^*_{ijkl} - \kappa^*_{sijkl}}{\sigma^*_{ijkl}}\right) \\ &= \int_{-\infty}^{\frac{\mu^*_{ijkl} - \kappa^*_{sijkl}}{\sigma^*_{ijkl}}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \end{aligned} \tag{2.8.13}$$

and for $(ijkl) \neq (ijkl)'$

$$P\left(Y_{tijkl} \geq s, Y_{t(ijkl)'} \geq s'\right) = \int_{-\infty}^{\frac{\mu^*_{ijkl} - \kappa^*_{sijkl}}{\sigma^*_{ijkl}}} \int_{-\infty}^{\frac{\mu^*_{(ijkl)'} - \kappa^*_{(sijkl)'}}{\sigma^*_{(ijkl)'}}} \phi_2\left(z^*_{ijkl}, z^*_{(ijkl)'}\right) dz^*_{ijkl} dz^*_{(ijkl)'} \tag{2.8.14}$$

with $z^*_{ijkl} = (Y^*_{ijkl} - \mu^*_{ij1l})/\sigma^*_{ijkl}$ and $z^*_{(ijkl)'} = (Y^*_{(ijkl)'} - \mu^*_{(ijkl)'})/\sigma^*_{(ijkl)'}$, and with ϕ_2 being the density of a bivariate standard normal distribution

$$\phi_2(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left\{-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right\} \tag{2.8.15}$$

with correlation coefficient ρ for the variables Y^*_{tijkl} and $Y^*_{t(ijkl)'}$. ρ corresponds to the polychoric correlation between Y_{tijkl} and $Y_{t(ijkl)'}$ that can be represented in terms of the model parameters as

$$\rho(z^*_{tijkl}, z^*_{t(ijkl)'}) = \rho(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = \frac{1}{\sigma^*_{ijkl}} Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) \frac{1}{\sigma^*_{(ijkl)'}} \tag{2.8.16}$$

with

$$Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = Cov(\pi_{tijkl}, \pi_{t(ijkl)'}) \tag{2.8.17}$$

2.9 Independence assumptions and testability

2.9.1 LS-Com GRM with conditional independence

In order to derive testable consequences of the LS-Com GRM, several independence assumptions have to be introduced. These assumptions define the LS-Com GRM with conditional independence. Note that classical assumptions of multilevel modeling are made, that is, the targets are assumed to be independently and randomly drawn from a set of targets and the interchangeable raters are assumed to be independently and randomly drawn from a set of interchangeable raters given a target. The following assumptions extend these independence assumptions.

Definition 2.3. (LS-Com GRM with conditional independence)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\pi_{\mathbf{rt}}$, $\pi_{\mathbf{t}}$, κ, α, $\mathbf{UM_{rt}}$, $\mathbf{CM_t}$, $\mathbf{M_t}$, $\lambda_{\mathbf{S}}$, $\lambda_{\mathbf{UM}}$, $\lambda_{\mathbf{CM}}$, $\lambda_{\mathbf{M}}\rangle$ be an LS-Com GRM with (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric variables. $\mathcal{M}$ is called LS-Com GRM with conditional independence if and only if for all y_{rtij2l}, $y_{tijkl} \in S_{ij}$ the following statements hold

1. *$(p_T, p_{TS_1}, \ldots, p_{TS_f})$-, $(p_T, p_{TS_1}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_2S_f})$- and $(p_T, p_{TS_1}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_f})$-conditional independence of the observed random variables Y_{tij1l}, Y_{rtij2l} and Y_{tijkl} $(k > 2)$.*

$$P\left(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}(Y_{rtij2l} = y_{rtij2l})\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\le e, k\ne 2}(Y_{tijkl} = y_{tijkl})\,\middle|\, p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_eS_f}\right)$$
$$= \prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P\big(Y_{tij1l} = y_{tij1l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}\big)$$
$$\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P\big(Y_{rtij2l} = y_{rtij2l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_2S_l}, \ldots, p_{R_2S_f}\big)$$
$$\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}\prod_{k=3}^{e} P\big(Y_{tijkl} = y_{tijkl} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f}\big) \quad (2.9.1)$$

2. *Situational conditional independence. For y_{rtij2l}, $y_{tijkl} \in S_{ij}$:*

$$P\big(Y_{tij1l} = y_{tij1l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}\big) = P\big(Y_{tij1l} = y_{tij1l} \mid p_T, p_{TS_l}\big) \quad (2.9.2)$$

$$P\big(Y_{rtij2l} = y_{rtij2l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_2S_l}, \ldots, p_{R_2S_f}\big) = P\big(Y_{rtij2l} = y_{rtij2l} \mid p_T, p_{TS_l}, p_R, p_{R_2S_l}\big) \quad (2.9.3)$$

$$P\big(Y_{tijkl} = y_{tijkl} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f}\big)$$
$$= P\big(Y_{tijkl} = y_{tijkl} \mid p_T, p_{TS_l}, p_{R_kS_l}\big) \qquad k > 2 \qquad (2.9.4)$$

3. *(p_T, p_{TS_l})-conditional regressive independence of the latent response variables π_{rtij2l}:*

$$\mathbb{E}\big(\pi_{rtij2l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f}\big)$$
$$= \mathbb{E}\big(\pi_{rtij2l} \mid p_T, p_{TS_l}\big) \qquad k > 2 \qquad (2.9.5)$$

Remarks. According to Equation (2.9.1), the response of a person on an item given the values on the variables $p_T, p_{TS_1}, \ldots, p_{TS_f}$ ($k = 1$; self-report), $p_T, p_{TS_1}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_2S_f}$ ($k = 2$; interchangeable informant report), or $p_T, p_{TS_1}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_f}$ ($k > 2$; structurally different informant report) is independent of all other item responses. The responses of all targets and raters on item i of construct j on measurement occasion l are independent given the projections p_T, p_{TS_l}, p_R and $p_{R_kS_l}$. That is, every dependency between the observed variables is determined entirely by the target, the rater, the target-situation, and the rater-situations that are given by the respective projections.
According to Equation (2.9.2), the probability that the variable Y_{tij1l} assumes a value $y_{tij1l} \in S_{ij}$ depends only on the target and its situation on the given measurement occasion l, but, given the former, not on situations on other measurement occasions l', $l' \neq l$. Analogously, according to Equation (2.9.3), the probability that the variable Y_{rtijkl} assumes a value $y_{rtijkl} \in S_{ij}$, given the target, the rater, the target- and the rater-situations on the given measurement occasion, is independent of the situations of the target or the rater on different measurement occasions l', $l' \neq l$. The same holds for the variables Y_{tijkl}, $k > 2$, by Equation (2.9.4).
Equation (2.9.5) states that given a target (p_T) and a target-situation (p_{TS_l}), the level-1 latent response variable π_{rtij2l} on measurement occasion l does not depend on target-situations on other measurement occasions or on the situations of other raters rating the same target (k >2).

The conditional independence assumptions given in Definition 2.3 imply consequences regarding the conditional and unconditional distributions of the observed variables Y_{tijkl} and Y_{rtij2l} as well as a specific covariance structure of the latent variables π_{tijkl} and π_{rtij2l} in the LS-Com GRM. Whether the restrictions imposed on the probability distributions of the response vectors and on the covariance structure by the conditional independence assumptions hold in empirical applications can be tested. That is, the conditional independence assumptions given in Definition 2.3 impose testable consequences on the covariance structure of the LS-Com GRM. These are derived in Section 2.10 and the following theorem.

Theorem 2.3. ***(LS-Com GRM with conditional independence)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \pi_{\mathbf{rt}}, \pi_{\mathbf{t}}, \kappa, \alpha, \mathbf{UM_{rt}}, \mathbf{CM_t}, \mathbf{M_t}, \lambda_{\mathbf{S}}, \lambda_{\mathbf{UM}}, \lambda_{\mathbf{CM}}, \lambda_{\mathbf{M}}\rangle$ be an LS-Com GRM of (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric variables with conditional independence. Then, for all

$j, j' \in J$, $i, i' \in I_j$, $k \in K$, $l, l' \in L$, and y_{rtij2l}, $y_{tijkl} \in S_{ij}$ it holds that:

$$P\Bigg(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\big(Y_{rtij2l}=y_{rtij2l}\big)\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\le e, k\neq 2}\big(Y_{tijkl}=y_{tijkl}\big)\,\Bigg|\,\pi_{t1111},\ldots,\pi_{tc_ddef},\pi_{rt1121},\ldots,\pi_{rtc_dd2f}\Bigg) \tag{2.9.6}$$
$$=\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}\prod_{k\le e, k\neq 2}P\big(Y_{tijkl}=y_{tijkl}\mid \pi_{tijkl}\big)\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P\big(Y_{rtij2l}=y_{rtij2l}\mid \pi_{rtij2l}\big)$$

Furthermore, it holds that:

$$P\Bigg(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\big(Y_{rtij2l}=y_{rtij2l}\big)\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\le e, k\neq 2}\big(Y_{tijkl}=y_{tijkl}\big)\,\Bigg|\,S_{t1111},\ldots,S_{tc_dd1f},M_{t131},\ldots,M_{tdef},CM_{t121},\ldots,CM_{td2f},UM_{rt121},\ldots,UM_{rtd2f}\Bigg)$$
$$=\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P\big(Y_{tij1l}=y_{tij1l}\mid S_{tij1l}\big)\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P\big(Y_{rtij2l}=y_{rtij2l}\mid S_{tij1l},CM_{tj2l},UM_{rtj2l}\big)$$
$$\prod_{l=1}^{f}\prod_{k=3}^{e}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P\big(Y_{tijkl}=y_{tijkl}\mid S_{tij1l},M_{tjkl}\big) \tag{2.9.7}$$

Remarks. Equation (2.9.6) follows from Equations (2.9.1) and (2.9.2) - (2.9.4). This is the case as the random variables π_{tij1l}, π_{tijkl}, $k > 2$, and π_{rtij2l} are (p_T, p_{TS_l})-, (p_T, p_{TS_l}, $p_{R_kS_l}$)-, and (p_T, p_{TS_l}, p_R, $p_{R_2S_l}$)-measurable functions, respectively. Similar arguments lead to Equation (2.9.7). A prove was given by Eid (1995, pp. 97-98) for a comparable model and is applicable to the present case.
According to Equation (2.9.6), all observed variables Y_{rtij2l} and Y_{tijkl} are independent given the latent response variables π_{rtij2l} and π_{tijkl}. Equation (2.9.6) implies that all associations between the observed variables are determined by the latent variables π_{tijkl} and π_{rtij2l} and their associations. According to Equation (2.9.7), the same holds with respect to the variables S_{tij1l}, M_{tjkl}, CM_{tj2l}, and UM_{rtj2l}.

2.9.2 LS-Com GRM in subpopulations

If the LS-Com GRM holds in a population, the model implies that it also holds in every subpopulation. That is, the item parameters α, λ, and κ have the same value in different subpopulations and the values on the latent variables π_{tijkl}, π_{rtij2l}, S_{tijkl}, UM_{rtij2l}, CM_{tij2l}, and M_{tijkl} remain the same when considering subpopulations, given that the parameterization and scaling of the latent variables is the same. This fact was proven by Eid (1995, pp. 94-96, 99) for a comparable model. The prove applies to the present model, too, and shall therefore not be repeated here.
Furthermore, if an LS-Com GRM with conditional independence holds in a population, the same conditional independence assumptions also hold in subpopulations. That is, the covariance structure

implied by the conditional independence assumptions has to hold in every subpopulation. For a prove for a comparable model see Eid (1995; also see Steyer, 1989). While the covariance structure has to be the same in every subpopulation, the values of the (non-zero) variances and covariances between the latent variables are allowed to vary between subpopulations.

2.10 Covariance structure

The LS-Com GRM with conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}. The following theorem introduces the covariances that are zero as a result of the conditional independence assumptions.

Theorem 2.4. ***(Testability)***

If $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}),\ \boldsymbol{\pi_{rt}},\ \boldsymbol{\pi_{t}},\ \kappa,\ \alpha,\ \mathbf{UM_{rt}},\ \mathbf{CM_{t}},\ \mathbf{M_{t}},\ \boldsymbol{\lambda_{S}},\ \boldsymbol{\lambda_{UM}},\ \boldsymbol{\lambda_{CM}},\ \boldsymbol{\lambda_{M}}\rangle$ is an LS-Com GRM of (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric variables with conditional independence, then, for all $j, j' \in J$, $i, i' \in I_j$, $k \in K$, and $l, l' \in L$, it holds that

1. *The latent state variables are uncorrelated with the latent method variables:*

$$Cov(S_{tij1l}, CM_{tj2l}) = 0 \tag{2.10.1}$$
$$Cov(S_{tij1l}, UM_{rtj'2l'}) = 0 \tag{2.10.2}$$
$$Cov(S_{tij1l}, M_{tjkl}) = 0 \tag{2.10.3}$$

2. *The unique method variables are uncorrelated with all of the level-2 method variables*

$$Cov(CM_{tj2l}, UM_{rtj'2l'}) = 0 \tag{2.10.4}$$
$$Cov(M_{tjkl}, UM_{rtj'2l'}) = 0 \tag{2.10.5}$$

Proofs. *Testability.*
The following proofs are based on Definitions 2.1 and 2.3 as well as general properties of residual variables. These properties are that residual variables are always uncorrelated with their regressors as well as with measurable functions of their regressors (Steyer & Nagel, 2017, p. 323).

2.10.1 By Equation (2.4.19)*, the common method factor CM_{tj2l} can be rewritten as $CM_{tj2l} = \frac{CM_{tij2l}}{\lambda_{CMij2l}}$, that is, it holds that: $Cov(S_{tij1l}, CM_{tj2l}) = 0 \iff Cov(S_{tij1l}, CM_{tij2l}) = 0$.*
CM_{tij2l} is defined as $CM_{tij2l} = S_{tij2l} - \mathbb{E}[S_{tij2l} \mid S_{tij1l}]$ by Equation (2.4.13)*. Hence, CM_{tij2l} is defined as a residual with respect to S_{tij1l}. As residuals are uncorrelated with their regressors, it follows that, for the same construct j and measurement occasion l, $Cov(S_{tij1l}, CM_{tj2l}) = 0$.*

2.10.2 By Equation (2.4.20)*, the unique method factor $UM_{rtj'2l'}$ can be rewritten as $UM_{rtj'2l'} = \frac{UM_{rti'j'2l'}}{\lambda_{UMi'j'2l'}}$, that is, it holds that: $Cov(S_{tij1l}, UM_{rtj'2l'}) = 0 \iff Cov(S_{tij1l}, UM_{rti'j'2l'}) = 0$.*
The latent state variable S_{tij1l} is given by $S_{tij1l} = \pi_{tij1l}$, which is a (p_T, p_{TS_l})-measurable function as it can be defined as $\pi_{tij1l} : \varphi_{tij1l}(p_T, p_{TS_l})$, with $\varphi_{tij1l} : \Omega_T \times \Omega_{TS_l} \rightarrow \mathbb{R}$ and (p_T, p_{TS_l}) :

$\Omega \to \Omega_T \times \Omega_{TS_l}$ (see remarks to Definition 2.1 in Section 2.4).
The variable $UM_{rti'j'2l'}$ is defined as $UM_{rti'j'2l'} = \pi_{rti'j'2l'} - \mathbb{E}[\pi_{rti'j'2l'} \mid p_T, p_{TS_{l'}}]$ by Equation 2.4.9. By conditional independence Assumption (2.9.5) *the expression $\mathbb{E}(\pi_{rtij2l} \mid p_T, p_{TS_{l'}})$ can be replaced by $\mathbb{E}(\pi_{rtij2l} \mid p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f})$. Therefore, it follows that $UM_{rti'j'2l'}$ is also a residual with respect to a $(p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f})$-measurable function, and thereby uncorrelated to the (p_T, p_{TS_l})-measurable function S_{tij1l}.*

2.10.3 By Equation (2.4.21)*, the method factor M_{tjkl} can be rewritten as $M_{tjkl} = \frac{M_{tijkl}}{\lambda_{Mijkl}}$, that is, it holds that: $Cov(S_{tij1l}, M_{tjkl}) = 0 \iff Cov(S_{tij1l}, M_{tijkl}) = 0$, for all $k > 2$.*
M_{tijkl}, $k > 2$, is defined as $M_{tijkl} = S_{tijkl} - \mathbb{E}[S_{tijkl} \mid S_{tij1l}]$ by Equation (2.4.14)*. Hence, M_{tijkl} is defined as a residual with respect to S_{tij1l}. As residuals are uncorrelated with their regressors, it follows that, for the same construct j and measurement occasion l, $Cov(S_{tij1l}, M_{tjkl}) = 0$.*

2.10.4 By Equations (2.4.19) *and* (2.4.20)*, it holds that:*
$Cov(CM_{tj2l}, UM_{rtj'2l'}) = 0 \iff Cov(CM_{tij2l}, UM_{rti'j'2l'}) = 0$.
CM_{tij2l} is defined as $CM_{tij2l} = S_{tij2l} - \mathbb{E}[S_{tij2l} \mid S_{tij1l}]$ by Equation (2.4.13)*, with S_{tij2l} given by $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$ by Equation 2.4.11. S_{tij2l} and $S_{tij1l} = \pi_{tij1l}$ are (p_T, p_{TS_l})-measurable functions, therefore CM_{tij2l} is a (p_T, p_{TS_l})-measurable function, too. $UM_{rti'j'2l'}$ on the other hand is a residual with respect to a (p_T, p_{TS_l})-measurable function by conditional independence Assumption* (2.9.5) *(see Proof 2.10.2). It follows that $Cov(CM_{tj2l}, UM_{rtj'2l'}) = 0$ for all $j, j' \in J$ and $l, l' \in L$.*

2.10.5 By Equations (2.4.20) *and* (2.4.21)*, it holds that:*
$Cov(M_{tjkl}, UM_{rtj'2l'}) = 0 \iff Cov(M_{tijkl}, UM_{rti'j'2l'}) = 0$.
M_{tijkl}, $k > 2$, is defined as $M_{tijkl} = S_{tijkl} - \mathbb{E}[S_{tijkl} \mid S_{tij1l}]$ by Equation (2.4.14)*. For $k > 2$, the variable $S_{tijkl} = \pi_{tijkl}$ is a $(p_T, p_{TS_l}, p_{R_kS_l})$- measurable function (see remarks to Definition 2.1 in Section 2.4), while $S_{tij1l} = \pi_{tij1l}$ is a (p_T, p_{TS_l})- measurable function. Hence, the variable M_{tijkl} is a $(p_T, p_{TS_l}, p_{R_kS_l})$-measurable function.*
The unique method variable $UM_{rti'j'2l'}$ is given by $UM_{rti'j'2l'} = \pi_{rti'j'2l'} - \mathbb{E}[\pi_{rti'j'2l'} \mid p_T, p_{TS_{l'}}]$. From conditional independence Assumption (2.9.5) *and the definition of the variable $UM_{rti'j'2l'}$ it follows that $UM_{rti'j'2l'}$ is a residual with respect to a $(p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f})$-measurable function ($k > 2$) and is therefore uncorrelated with the $(p_T, p_{TS_l}, p_{R_kS_l})$-measurable function M_{tijkl} ($k > 2$).*

The LS-Com GRM with conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}, including the zero-correlations specified in Theorem 2.4. Whether this covariance structure holds in empirical applications can be tested based on the covariance structure of the variables Y^*_{tijkl} and Y^*_{rtij2l}, as defined in Section 2.8, with SEMs for ordinal observed variables. Therefore, the covariance structure of the variables Y^*_{tijkl} and Y^*_{rtij2l} will be derived in the following.
According to Equation (2.9.6), all observed variables Y_{rtij2l} and Y_{tijkl} are independent given the latent response variables π_{rtij2l} and π_{tijkl}. Equation (2.9.6) implies that all associations between the observed variables are determined by the latent variables π_{tijkl} and π_{rtij2l} and their associations. In the factor analytical representation, the associations between the observed variables Y_{rtij2l} and Y_{tijkl} are explained by the associations between the variables Y^*_{rtij2l} and Y^*_{tijkl}, which again are determined by the associations between the variables π_{rtij2l} and π_{tijkl}. If the variables Y_{rtij2l} and Y_{tijkl} are conditionally independent given the variables π_{rtij2l} and π_{tijkl}, the covariances between the variables Y^*_{rtij2l}

and Y^*_{tijkl} have to be zero given π_{rtij2l} and π_{tijkl}. That is, for the variables $Y^*_{tijkl} = \pi_{tijkl} + \varepsilon^*_{tijkl}$ and $Y^*_{rtij2l} = \pi_{rtij2l} + \varepsilon^*_{rtij2l}$ it has to hold that, for all $(i,j,k,l) \neq (i,j,k,l)'$:

$$Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = Cov(\pi_{tijkl}, \pi_{t(ijkl)'}) \tag{2.10.6}$$

$$Cov(Y^*_{rtij2l}, Y^*_{rt(ij2l)'}) = Cov(\pi_{rtij2l}, \pi_{rt(ij2l)'}) \tag{2.10.7}$$

and

$$Cov(Y^*_{rtij2l}, Y^*_{t(ijkl)'}) = Cov(\pi_{rtij2l}, \pi_{t(ijkl)'}) \tag{2.10.8}$$

This is the case as the variables Y^*_{rtij2l} and Y^*_{tijkl} cannot be defined on the probability space given by the random experiment and can therefore not be observed. The covariances between the Y^*_{rtij2l} and Y^*_{tijkl} cannot be computed but only estimated based on the associations between the variables Y_{rtij2l} and Y_{tijkl}. The observable variables Y_{rtij2l} and Y_{tijkl}, however, are independent given the latent response variables π_{rtij2l} and π_{tijkl} (by Equation 2.9.6). Consequently, there cannot be any associations between the Y^*_{rtij2l} and Y^*_{tijkl} variables that are not determined by the associations between the variables π_{rtij2l} and π_{tijkl} (see Eid, 1995). As

$$\begin{aligned} Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) &= Cov(\pi_{tijkl} + \varepsilon^*_{tijkl}, \pi_{t(ijkl)'} + \varepsilon^*_{t(ijkl)'}) \\ &= Cov(\pi_{tijkl}, \pi_{t(ijkl)'}) + Cov(\pi_{tijkl}, \varepsilon^*_{t(ijkl)'}) \\ &\quad + Cov(\varepsilon^*_{tijkl}, \pi_{t(ijkl)'}) + Cov(\varepsilon^*_{tijkl}, \varepsilon^*_{t(ijkl)'}) \end{aligned} \tag{2.10.9}$$

it has to hold that the $Cov(\varepsilon^*_{tijkl}, \pi_{t(ijkl)'}) = 0$ and $Cov(\varepsilon^*_{tijkl}, \varepsilon^*_{t(ijkl)'}) = 0$ for all $i, i' \in I_j$, $j, j' \in J$, $k, k' \in K$, and $l, l' \in L$. The same applies to the Y^*_{rtij2l} and their residuals and the combination of Y^*_{rtij2l} and Y^*_{tijkl} and their residuals. As the residuals ε^*_{tijkl} and ε^*_{rtij2l} have to be uncorrelated with all π_{tijkl} and π_{rtij2l}, the residual variables are as well uncorrelated with all latent state variables S_{tijkl}, unique method factors UM_{rtj2l}, common method factors CM_{tj2l} and method factors M_{tjkl}. This fact also follows from Equation (2.9.7), which states that all associations between the observed variables Y_{rtij2l} and Y_{tijkl} are determined by the corresponding latent variables S_{tij1l}, UM_{rtj2l}, CM_{tj2l} and M_{tjkl}.
The zero-correlations of the error variables with all other error variables and latent variables of the LS-Com GRM combined with the covariance structure of the latent response variables π_{tijkl} and π_{rtij2l} define the covariance structure of the variables Y^*_{rtij2l} and Y^*_{tijkl}. This covariance structure equals the covariance structure of the latent variables in the LS-Com model for continuous indicator variables derived by Koch (2013), with one exception. While the variance of the error variables is a variable that is free to vary and is estimated in the SEM with continuous indicator variables, this variance is fixed to one for ε^*_{tijkl} and ε^*_{rtij2l} for all $j \in J$, $i \in I_j$, $k \in K$, and $l \in L$ in the LS-Com GRM. This restriction guarantees the equivalence of the LS-Com GRM and the factor analytical representation of the model derived in Section 2.8.
In a nutshell, the total covariance matrix Σ_T of an LS-Com GRM with conditional independence can be partitioned, just as in the continuous case, into a within and a between covariance matrix and can be represented as

$$\Sigma_T = \Lambda_B \Phi_B \Lambda'_B + \Theta_B + \Lambda_W \Phi_W \Lambda'_W + \Theta_W \tag{2.10.10}$$

where Λ_B and Λ_W refer to the factor loading matrices of the between- and within-level factors, respectively, Φ_B and Φ_W refer to the variance-covariance matrices of the between and within latent variables, respectively, and Θ_B and Θ_W are the between- and within-level diagonal residual variance-covariance matrices. For a detailed illustration of these covariance matrices and their elements see

Koch (2013, pp. 40-45), where all non-zero elements $Var(E_{rtij2l})$ and $Var(E_{tijkl})$ in the matrices Θ_B and Θ_W have to be replaced by 1.
For a description of the interpretation of all non-zero covariances and correlations in the LS-Com model see Section 2.3 or refer to Koch (2013, pp. 46-50).

2.11 Variance decompositions

Based on the definition of the LS-Com GRM, the latent response variables π_{tijkl} and π_{rtij2l} can be additively decomposed into different variance components. From Definition 2.1 and Theorem 2.1 it follows that the general measurement equations for the latent response variables in an LS-Com GRM of $(CM_{tij2l}, UM_{rtij2l}, M_{tijkl})$-congeneric variables are given by:

$$\pi_{tij1l} = S_{tij1l} \tag{2.11.1}$$
$$\pi_{rtij2l} = \lambda_{Sij2l}S_{tij1l} + \lambda_{CMij2l}CM_{tj2l} + \lambda_{UMij2l}UM_{rtj2l} \tag{2.11.2}$$
$$\pi_{tijkl} = \lambda_{Sijkl}S_{tij1l} + \lambda_{Mijkl}M_{tjkl} \qquad k > 2 \tag{2.11.3}$$

As the latent method variables are defined as latent residual variables, they are uncorrelated with their respective regressors. That is, due to the zero-covariances given in Equations (2.10.1) - (2.10.5), the different variance components can be separated. The variances of the latent response variables can therefore be additively decomposed as:

$$Var(\pi_{tij1l}) = Var(S_{tij1l}) \tag{2.11.4}$$
$$Var(\pi_{rtij2l}) = \lambda^2_{Sij2l}Var(S_{tij1l}) + \lambda^2_{CMij2l}Var(CM_{tj2l}) + \lambda^2_{UMij2l}Var(UM_{rtj2l}) \tag{2.11.5}$$
$$Var(\pi_{tijkl}) = \lambda^2_{Sijkl}Var(S_{tij1l}) + \lambda^2_{Mijkl}Var(M_{tjkl}) \qquad k > 2 \tag{2.11.6}$$

Table 2.2: Definition of the consistency and different method specificity coefficients in the LS-Com GRM.

Consistency and method specificity coefficients			
Coefficient	Level	Method	Definition
Consistency	Target	Structurally different	$Con(\pi_{tijkl}) = \frac{\lambda^2_{S_{ijkl}}Var(S_{tij1l})}{\lambda^2_{S_{ijkl}}Var(S_{tij1l})+\lambda^2_{M_{ijkl}}Var(M_{tjkl})}$
	Target	Interchangeable	$Con(\pi_{tij2l}) = \frac{\lambda^2_{S_{ij2l}}Var(S_{tij1l})}{\lambda^2_{S_{ij2l}}Var(S_{tij1l})+\lambda^2_{CM_{ij2l}}Var(CM_{tj2l})}$
	Rater	Interchangeable	$Con(\pi_{rtij2l}) = \frac{\lambda^2_{S_{ij2l}}Var(S_{tij1l})}{\lambda^2_{S_{ij2l}}Var(S_{tij1l})+\lambda^2_{CM_{ij2l}}Var(CM_{tj2l})+\lambda^2_{UM_{ij2l}}Var(UM_{rtj2l})}$
Method specificity	Target	Structurally different	$MS(\pi_{tijkl}) = \frac{\lambda^2_{M_{ijkl}}Var(M_{jkl})}{Var(\pi_{tijkl})}$
Common method specificity	Target	Interchangeable	$CMS(\pi_{rtij2l}) = \frac{\lambda^2_{CM_{ij2l}}Var(CM_{j2l})}{Var(\pi_{rtij2l})}$
Unique method specificity	Rater	Interchangeable	$UMS(\pi_{rtij2l}) = \frac{\lambda^2_{UM_{ij2l}}Var(UM_{j2l})}{Var(\pi_{rtij2l})}$
Reliability	Target	Structurally different	$Rel(\pi_{tijkl}) = \frac{Var(\pi_{tijkl})}{Var(\pi_{tijkl})+1}$
	Rater	Interchangeable	$Rel(\pi_{rtij2l}) = \frac{Var(\pi_{rtij2l})}{Var(\pi_{rtij2l})+1}$
ICC	Rater	Interchangeable	$ICC(\pi_{rtij2l}) = \frac{\lambda^2_{S_{ij2l}}Var(S_{tij1l})+\lambda^2_{CM_{ij2l}}Var(CM_{j2l})}{Var(\pi_{rtij2l})}$

Note. Con: Consistency; MS: Method specificity; CMS: Common method specificity; UMS: Unique method specificity; ICC: intra-class correlation coefficient.

Then, analogous to the LS-Com model with continuous indicators, different variance components for the non-reference method indicators can be defined. Definitions of the variance coefficients are given in Table 2.2. These variance components correspond to the variance components introduced by Koch (2013), with the only difference that they are defined based on the latent response variables π_{tijkl} and π_{rtij2l}. They can be meaningfully interpreted, as they are invariant under admissible transformations, as shown in Section 2.5. For interpretations of the coefficients see Section 2.3 or Koch (2013).

2.12 Mean structure

The following theorem clarifies the mean structure of the latent variables in the LS-Com GRM. The mean structure of the latent variables is of interest in research questions investigating mean changes in the latent states over time. Furthermore, the mean structure is needed to derive identification conditions in Section 2.13.

Theorem 2.5. ***(Mean Structure)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \pi_{\mathbf{rt}}, \pi_{\mathbf{t}}, \kappa, \alpha, \mathbf{UM_{rt}}, \mathbf{CM_t}, \mathbf{M_t}, \lambda_{\mathbf{S}}, \lambda_{\mathbf{UM}}, \lambda_{\mathbf{CM}}, \lambda_{\mathbf{M}}\rangle$ *be an LS-Com GRM of* (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})*-congeneric variables with conditional independence. Without loss of generality the first method (k=1) is chosen as reference method and the second method (k=2) as the set of interchangeable methods. Then, for all* $j \in J$, $i \in I_j$, $k \in K$, *and* $l \in L$ *it holds that*

$$\mathbb{E}(UM_{rtj2l}) = 0 \tag{2.12.1}$$
$$\mathbb{E}(CM_{tj2l}) = 0 \tag{2.12.2}$$
$$\mathbb{E}(M_{tjkl}) = 0 \qquad k > 2 \tag{2.12.3}$$

$$\mathbb{E}(\pi_{tsijkl}) = \mathbb{E}(\pi_{tijkl}) - \kappa_{sijkl} \qquad k \neq 2 \tag{2.12.4}$$
$$\mathbb{E}(\pi_{rtsij2l}) = \mathbb{E}(\pi_{rtij2l}) - \kappa_{sij2l} \tag{2.12.5}$$

$$\mathbb{E}(\pi_{tij1l}) = \mathbb{E}(S_{tij1l}) \tag{2.12.6}$$
$$\mathbb{E}(\pi_{tijkl}) = \alpha_{ijkl} + \lambda_{Sijkl}\,\mathbb{E}(S_{tij1l}) \qquad k > 2 \tag{2.12.7}$$
$$\mathbb{E}(\pi_{rtij2l}) = \alpha_{ij2l} + \lambda_{Sij2l}\,\mathbb{E}(S_{tij1l}) \tag{2.12.8}$$

and in LS-Com GRMs defined with common latent state factors:

$$\mathbb{E}(\pi_{tij1l}) = \lambda_{Sij1l}\big(\mathbb{E}(S_{tj1l}) + \delta_{ij1l}\big) \tag{2.12.9}$$

Proofs. *Mean Structure.*
Equations (2.12.1) - (2.12.3) *follow directly from the definition of the latent method variables as residual variables in Definition 2.1 and the fact that residual variables have an expectation of zero (Steyer & Nagel, 2017, p. 323). Equations* (2.12.4) - (2.12.9) *follow directly from the definitions of the latent response variables* π_{tsijkl}, $\pi_{rtsij2l}$, π_{tijkl}, *and* π_{rtij2l} *given in Definition 2.1 and Equation 2.6.1 as well as from Equations* (2.12.1) - (2.12.3).

Remarks. Equations (2.12.6) - (2.12.8) show that the expected value of the common latent response variables π_{tijkl}, $k > 2$, and π_{rtij2l} equal the expectation of the latent state factors S_{tij1l} if and only if $\alpha_{ijkl} = 0$ and $\lambda_{Sijkl} = 1$. For models defined with common latent state factors, the expectation of the latent response variables π_{tij1l} equal the expectation of the latent state factor S_{tj1l} if and only if $\delta_{ij1l} = 0$ and $\lambda_{Sij1l} = 1$. The effect of different identification variants and parameter invariance settings on the interpretation of latent state means and latent state mean differences is discussed in Section 2.13.

2.13 Identifiability

The identification problem deals with the question whether a unique solution exists for each of the model parameters that are to be estimated (Bollen, 1989). A model is identified if every parameter in the model can be represented as a function of known quantities, that is, information contained in the data and parameters that were constrained (e.g. for scaling the latent variables), and if there is only one mathematical solution for each parameter. In order to assign a scale to each latent factor, either one factor loading per factor or the variance of the latent factor has to be fixed to a value larger than 0 (typically 1; Bollen, 1989). As in longitudinal SEMs the interest often lies in investigating the change or stability of factor variances over time, it is preferable to choose the scaling that fixes one of the loading parameters (Geiser, 2008; Koch, 2013). As shown in Theorem 2.2, the latent response variables π_{rtijkl} and π_{tijkl}, their respective threshold variables κ_{sijkl}, and the variables α_{ijkl} are uniquely defined only up to translations. Consequently, the parameters α_{ijkl} and κ_{sijkl} are not separately identifiable. The same holds for the parameters κ_{sijkl} and δ_{ij1l} defined in Section 2.6 for the case of models with common latent state factors. Therefore, without loss of generality, the variables δ_{ij1l} are set to zero for all i, j, and l. Furthermore, recall that all latent method factors have an expectation of zero by definition.
In Equation (2.10.10) the total covariance matrix of an LS-Com GRM was represented as

$$\Sigma_T = \Lambda_B \Phi_B \Lambda_B' + \Theta_B + \Lambda_W \Phi_W \Lambda_W' + \Theta_W$$

where all non-zero elements in the residual variance-covariance matrices Θ_B and Θ_W are equal to 1. Theorem 2.6 then gives identification conditions for the LS-Com GRM parameters.

Theorem 2.6. ***(Identification of the LS-Com GRM)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}),\ \pi_{\mathbf{rt}},\ \pi_{\mathbf{t}},\ \kappa,\ \mathbf{UM_{rt}},\ \mathbf{CM_t},\ \mathbf{M_t},\ \lambda_{\mathbf{S}},\ \lambda_{\mathbf{UM}},\ \lambda_{\mathbf{CM}},\ \lambda_{\mathbf{M}}\rangle$ be an LS-Com GRM of (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric as well as S_{tij1l}-congeneric variables with conditional independence. The parameters in the matrices Λ_B, Λ_W, Φ_B, and Φ_W as well as the threshold parameters κ and the expectations of the latent state variables S_{tj1l} are identified if all δ_{ij1l} are set to zero and

(a) $i_j \geq 3$ for all j, $j \geq 1$, $k \geq 2$, $l \geq 1$, and:

1. *either one factor loading λ_{Sij1l}, λ_{CMij2l}, λ_{UMij2l}, λ_{Mijkl} for each factor S_{tj1l}, CM_{rtj2l}, UM_{rtj2l}, and M_{tjkl}, or the variance of the factors is set to any real value larger than 0, and*

2. *the mean of the latent state factors S_{tj11} belonging to the first occasion of measurement $l = 1$ is set to any real value for all j or one threshold κ_{sij11} of one indicator Y_{tij11} per latent state factor belonging to the first occasion of measurement $l = 1$ is fixed at any real value for all j, and*
3. *the mean of all remaining latent state factors S_{tj1l}, $l > 1$, is set to any real value or one threshold of one indicator Y_{tij1l} per latent state factor on measurement occasions $l > 1$ is fixed to any real value or is constrained to be invariant over measurement occasions for all j, that is κ_{sij1l} is fixed or is set as $\kappa_{sij1l} = \kappa_{sij11}$ for a chosen value of s and i, and for all j and l.*
4. *either α_{ijk1} or one κ_{sijk1} is set to any real value for all i, j, and $k > 1$ at $l = 1$, and either α_{ijkl} or one κ_{sijkl} is set to any real value or set invariant over measurement occasions for all i, j, $k > 1$, and all $l > 1$.*

(b) $i_j \geq 2$ with $i_j = 2$ for some j, $j \geq 1$, $k \geq 2$, $l \geq 2$, and:

conditions (1) - (4) in (a) hold and Φ_B as well as Φ_W contain substantive (permissible) intercorrelations among the latent variables for the respective j with $i_j = 2$.

Condition (1) of (a) and (b) is identical to the continuous indicator case and identifies the parameters of the LS-Com GRM covariance structure (that is, the parameters in the matrices Λ_B, Λ_W, Φ_B, and Φ_W), given the polychoric correlations between the variables Y^*_{rtij2l} and Y^*_{tijkl} (or π_{rtij2l} and π_{tijkl}). The identification of this part of the model was shown by Geiser (2008) and Koch (2013) for the continuous indicator model and is applicable to the LS-Com GRM, too.

In an LS-Com GRM with indicator-specific latent state variables, the model is identified under the conditions given in (a) and (b) if Φ_B contains substantive correlations between the indicator-specific state variables or if $k \geq 3$. Note that in the case of indicator-specific latent state variables, the loading parameters for the loading of the reference method ($k = 1$) indicators on the latent state variables S_{tij1l} already correspond to 1 by definition (see Equation 2.4.10 in Definition 2.1).

The indicators for which the loadings are set to unity by Condition (1) are referred to as reference indicators in the following.

Conditions (2) - (4) are needed for the identification of the means μ_{ijkl} of the latent response variables and thereby the means of the latent state variables as well as the threshold variables κ_{sijkl}.

Millsap and Yun-Tein (2004) provided minimal identification conditions for the multiple-population case of factor analyses of ordered-categorical measures, as it applies to longitudinal ordered categorical CFAs, too. The identification conditions for the LS-Com GRM given in Theorem 2.6 deviate from the conditions given by Millsap and Yun-Tein (2004) in that only one threshold is required to be invariant over measurement occasions for only one indicator of the factors S_{tij1l} (instead of two thresholds for the reference indicators and one threshold for every remaining indicator). This is due to the restriction that the residual variances have to equal one on all occasions in the LS-Com GRM to ensure equivalence to a graded response model. Fixing two thresholds for the reference indicator allows to freely estimate residual variances for $l > 1$ and would lead to an over-identified model when leaving the residual variances restricted to a value of 1. Without providing a formal proof (see Millsap & Yun-Tein, 2004), identification of the model parameters under the conditions provided in Theorem 2.6 shall be roughly delineated in the following. For the sake of clarity and due to space restrictions, this will be restricted to a model with common latent state factors. Identification for the case of indicator-specific latent state factors works along the same lines. The demonstration of the identification of the model parameters is based on the factor analytical representation of the LS-Com GRM

presented in Section 2.8. As the parameters of the factor analytical representation with a standardization of $\theta_{ijkl} = 1$ are transformations of the parameter estimates obtained with the standardization of $\sigma^*_{ijkl} = 1$, identification of the model under the first standardization implies the identification of the model under the second standardization. For simplicity, identification is shown starting with a standardization of $\sigma^*_{ijkl} = 1$. The parameters of the LS-Com GRM under the restriction $\theta_{ijkl} = 1$ can be easily obtained from the former.
Identification conditions can be derived from the univariate and bivariate marginal probability expressions in the factor analytical representation given in Equations (2.8.13) and (2.8.14) as well as from the expressions for the latent correlations and covariances in Equations (2.8.16) and (2.8.17).

Let the standardized thresholds of a variable Y^*_{tijkl} or Y^*_{rtij2l} be denoted by

$$z_{sijkl} := \frac{\mu^*_{ijkl} - \kappa^*_{sijkl}}{\sigma^*_{ijkl}} \tag{2.13.1}$$

These standardized thresholds z_{sijkl} are identified for all s, i, j, k, and l as the respective z-scores of the univariate standard normal distribution by Equation (2.8.13). Given the identification of the standardized thresholds z_{sijkl}, the polychoric correlation ρ of two variables Y^*_{tijkl} and $Y^*_{t(ijkl)'}$ is identified by Equation (2.8.14), i.e.,

$$P\Big(Y_{tijkl} \geq s, Y_{t(ijkl)'} \geq s'\Big) = \int_{-\infty}^{\frac{\mu^*_{ijkl} - \kappa^*_{sijkl}}{\sigma^*_{ijkl}}} \int_{-\infty}^{\frac{\mu^*_{(ijkl)'} - \kappa^*_{(sijkl)'}}{\sigma^*_{(ijkl)'}}} \phi_2\big(z^*_{ijkl}, z^*_{(ijkl)'}\big) dz^*_{ijkl} dz^*_{(ijkl)'}$$

with ϕ_2 being the density of a bivariate standard normal distribution with correlation coefficient ρ, z^*_{ijkl} and $z^*_{(ijkl)'}$ being the standardized latent response variables, i.e., $z^*_{ijkl} = (Y^*_{ijkl} - \mu^*_{ij1l})/\sigma^*_{ijkl}$ and $z^*_{(ijkl)'} = (Y^*_{(ijkl)'} - \mu^*_{(ijkl)'})/\sigma^*_{(ijkl)'}$, and $\rho(z^*_{tijkl}, z^*_{t(ijkl)'}) = \rho(Y^*_{tijkl}, Y^*_{t(ijkl)'})$.
Consider the standardization $\sigma^*_{ijkl} = 1$. Then the correlation

$$\rho(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = \frac{1}{\sigma^*_{ijkl}} Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) \frac{1}{\sigma^*_{(ijkl)'}}$$

corresponds to the respective covariance

$$Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = Cov(\pi_{tijkl}, \pi_{t(ijkl)'})$$

In the following, it is assumed that the covariance structure of the latent response variables $\pi_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$ (i.e., the expectations of the within-level latent response variables over clusters) is available. This is the standard assumption that cluster-level random effects are (in principal) estimable in multilevel models for ordinal data. The estimation of these target-level random effects in multilevel IRT models are for instance described in Rabe-Hesketh, Skrondal, and Pickles (2005) or, using Bayesian estimation with a Gibbs sampler, in Fox and Glas (2001). Then, consider, for instance, the covariances for the different latent response variables for the same measurement occasion l and construct j, with $i = i'$ or $i \neq i'$. These correspond to:

$$Cov(\pi_{tijkl}, \pi_{ti'jk'l}) = \lambda_{Sijkl}\lambda_{Si'jk'l}Var(S_{tj1l}), \qquad k = k' = 1,\ or\ k = 1\ and\ k' \neq 1 \tag{2.13.2}$$

$$Cov(\pi_{tij2l}, \pi_{ti'j2l}) = \lambda_{Sij2l}\lambda_{Si'j2l}Var(S_{tj1l}) + \lambda_{CMij2l}\lambda_{CMi'j2l}Var(CM_{tj2l}) \tag{2.13.3}$$

$$Cov(\pi_{tijkl}, \pi_{ti'jkl}) = \lambda_{Sijkl}\lambda_{Si'jkl}Var(S_{tj1l}) + \lambda_{Mijkl}\lambda_{Mi'jkl}Var(M_{tjkl}), \qquad k > 2 \tag{2.13.4}$$

$$Cov(\pi_{tij2l}, \pi_{ti'jkl}) = \lambda_{Sij2l}\lambda_{Si'jkl}Var(S_{tj1l}) + \lambda_{CMij2l}\lambda_{Mi'jkl}Cov(CM_{tj2l}, M_{tjkl}), \qquad k > 2 \tag{2.13.5}$$

$$Cov(\pi_{rtij2l}, \pi_{rti'j2l}) = \lambda_{Sij2l}\lambda_{Si'j2l}Var(S_{tj1l}) + \lambda_{CMij2l}\lambda_{CMi'j2l}Var(CM_{tj2l}) + \lambda_{UMij2l}\lambda_{UMi'j2l}Var(UM_{rtj2l}) \tag{2.13.6}$$

Hence, the covariances between the latent response variables π_{tijkl} as well as π_{rtij2l} belonging to the same measurement occasion l are functions of the loading parameters and variances of the latent state and method factors. The covariances between the latent response variables π_{tijkl} as well as π_{rtij2l} belonging to different measurement occasions $l \neq l'$, on the other hand, are functions of the loading parameters and covariances between the latent state and method factors, respectively. From the above equations for the covariances it can be seen that Condition (1) in (a) and (b) identifies the parameters of the LS-Com GRM covariance structure, that is, the loading parameters in the matrices Λ_B and Λ_W as well as the variances and covariances in the matrices Φ_B and Φ_W. The identification of these parameters by the constraints set in Condition (1) is analogous to the continuous indicator case and is derived in detail in Geiser (2008) and complemented for the LS-Com model by Koch (2013). Given the identification of the variances of the latent state and method factors as well as the loading parameters, the variances Ψ_{ijkl} of the latent response variables π_{tijkl} and π_{rtij2l} are identified. The estimated loading parameters can then be rescaled using Equation 2.8.12 and the variances σ^*_{ijkl} can be calculated to yield parameters corresponding to the parameterization $\theta_{ijkl} = 1$.

Given the estimates of the variances σ^*_{ijkl} as well as the standardized thresholds z_{sijkl}, the expression $z_{sijkl} := \frac{\mu^*_{ijkl} - \kappa_{sijkl}}{\sigma^*_{ijkl}}$ contains two unknowns: the thresholds κ_{sijkl} as well as the means μ^*_{ijkl} of the latent response variables π_{tijkl} and π_{rtij2l}. Setting $\alpha_{ijkl} = 0$ and $\delta_{ij1l} = 0$ for all i, j, k and l, the expectation of the latent response variables π_{tijkl} and π_{rtij2l} are given by

$$\mu^*_{ijkl} = \mathbb{E}(\pi_{tijkl}) = \lambda_{Sijkl}\ \mathbb{E}(S_{tj1l}) \qquad k \neq 2 \tag{2.13.7}$$

$$\mu^*_{ij2l} = \mathbb{E}(\pi_{rtij2l}) = \lambda_{Sij2l}\ \mathbb{E}(S_{tj1l}) \tag{2.13.8}$$

according to Equations (2.12.7) and (2.12.8). Using the first variant of identification constraint (2), setting the expectation of S_{tj1l} to zero leads to $\mu^*_{ijkl} = \mathbb{E}(\pi_{tijkl}) = \mathbb{E}(\pi_{rtij2l}) = 0$. Then, the threshold parameters κ_{sijkl} are identified for all s, i, j, and k, for $l = 1$.

Now consider the variant where one threshold κ_{sijkl} for a selected indicator is fixed to any real value on measurement occasion $l = 1$, instead of standardizing to $\mathbb{E}(S_{tj1l}) = 0$. That is, κ_{sijk1} is fixed to any real value for a selected s, i and k for every j. Then, the means of the latent response variables at $l = 1$ are identified as:

$$\mu^*_{ijk1} = z_{sijk1}\sigma^*_{ijk1} + \kappa_{sijk1} \tag{2.13.9}$$

As the loading parameters λ_{Sijkl} are identified as described above, the expectation of the latent state variable S_{tj11} is then given by

$$\mathbb{E}(S_{tj11}) = \frac{\mu^*_{ijk1}}{\lambda_{Sijk1}} \tag{2.13.10}$$

Given the estimate $\mathbb{E}(S_{tj11})$ as well as the loading parameters, the means μ^*_{ijk1} of all other i and k, and thereby all other threshold parameters for all indicators loading on the same latent state variable

S_{tijk1} at the first measurement occasion $l = 1$ are identified by Equation (2.13.1). Then, fixing either the expectation of the latent state variable $\mathbb{E}(S_{tj1l}) = 0$, setting one threshold to any real value for a chosen s, i, and k for every j, or constraining one threshold for a chosen s, i, and k to be invariant over time for every j and $l > 1$, that is, $\kappa_{sijkl} = \kappa_{sijk1}$ for every j and a chosen value of s, i, and k, identifies all thresholds as well as expectations of the latent state variables S_{tj1l}, $l > 1$. Therefore, identification constraint (3) identifies all remaining parameters κ_{sijkl} and $\mathbb{E}(S_{tj1l})$ for all $l > 1$.

Although the means of the latent state variables S_{tij1l} (or S_{j1l} in the case of common latent state variables) are identifiable for all measurement occasions l, as demonstrated in Theorem 2.6, these may vary depending on the threshold that is fixed and the value of the loading λ_{Sijkl} of the respective indicator for which a threshold was constrained. That is, mean differences between latent state variables on different measurement occasions are not invariant under different standardization variants and therefore not necessarily comparable between measurement occasions. The interpretation of latent state means $\mathbb{E}(S_{tij1l})$ hence varies for different identification constraints and, for comparisons between different measurement occasions, the choice of identification constraints as well as invariance settings have to be taken into account (see Section 2.14).

2.14 Measurement invariance over time

Measurement invariance (MI) deals with the question whether a measure assesses the same latent construct in the same way in different groups or on different measurement occasions (Meredith & Teresi, 2006). MI is a prerequisite for the interpretation of test score differences between groups or time points as differences in the underlying constructs. The definition of MI is based on conditional distributions of the manifest variables given the underlying latent factors (Meredith, 1993). For ordered categorical observations, the concept of MI is based on the relationship between response probabilities and the underlying latent constructs (Meredith & Teresi, 2006). That is the case as MI refers to the observed measures, which are linked to the factor model only indirectly via the probabilities of observing a response in a certain category given the underlying latent variables. In the normal ogive GRM, MI thus holds if the probability functions for observing a response in category s or higher given the latent variables are invariant across groups or measurement occasions. This corresponds to the parameters determining the shape and position of the conditional probability curves being equal across groups. In accordance with a definition by Millsap and Yun-Tein (2004), MI for the ordered categorical measures Y_{tijkl} and Y_{rtij2l} with respect to S_{tij1l}, CM_{tj2l}, UM_{rtj2l}, and M_{tjkl} and the different occasions of measurement l, holds if

$$P(Y_{tij1l} = s \mid S_{tij1l}) = P(Y_{tij1l'} = s \mid S_{tij1l'}) \tag{2.14.1}$$

$$P(Y_{rtij2l} = s \mid S_{tij1l}, CM_{tj2l}, UM_{rtj2l}) = P(Y_{rtij2l'} = s \mid S_{tij1l'}, CM_{tj2l'}, UM_{rtj2l'}) \tag{2.14.2}$$

$$P(Y_{tijkl} = s \mid S_{tij1l}, M_{tjkl}) = P(Y_{tijkl'} = s \mid S_{tij1l'}, M_{tjkl'}), \qquad k > 2 \tag{2.14.3}$$

for all $l, l' \in L$, and all i, j, k and $s \in S_{ij}$. MI in the longitudinal LS-Com GRM would thereby be given if the conditional probabilities of the outcomes for Y_{tijkl} and Y_{rtij2l} given the latent variables S_{tij1l}, CM_{tj2l}, UM_{rtj2l} and M_{tjkl} are independent of the measurement occasion. For Conditions (2.14.1) - (2.14.3) to hold, not only the conditional distributions of Y^*_{tijkl} and Y^*_{rtij2l} given S_{tij1l}, CM_{tj2l}, UM_{rtj2l}, and M_{tjkl} have to be invariant over measurement occasions, but also the relationship between the observed categories and the underlying Y^*_{tijkl} and Y^*_{rtij2l} have to be the same. The former corresponds to strict factorial invariance for the latent variables Y^*_{tijkl} and Y^*_{rtij2l} as defined by Meredith (1993), while the latter corresponds to invariant threshold parameters over measurement occasions. Note that in the factor-analytical representation of the LS-Com GRM, the residual variances are fixed to 1 on all measurement occasions and are therefore invariant by definition.

As stated by Meredith and Teresi (2006), in the case of IRT models, no form of factorial invariance other than strict factorial invariance will be sufficient to establish MI. Definition 2.4 defines specific conditions for MI for the LS-Com GRM.

Definition 2.4. (LS-Com GRM with MI)

$\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}),\ \pi_{\mathbf{rt}},\ \pi_{\mathbf{t}},\ \kappa,\ \alpha,\ \mathbf{UM_{rt}},\ \mathbf{CM_t},\ \mathbf{M_t},\ \lambda_{\mathbf{S}},\ \lambda_{\mathbf{UM}},\ \lambda_{\mathbf{CM}},\ \lambda_{\mathbf{M}}\rangle$ is called LS-Com GRM of (CM_{tij2l}, UM_{rtij2l}, M_{tijkl})-congeneric variables with conditional independence and measurement invariance if and only if Definition 2.1 and 2.3 and Theorem 2.1 apply, and for all $i \in I_j$, $j \in J$, $k \in K$, $s \in S_{ij}$ and for l, $l' \in L$ the following statements hold:

$$\kappa_{sijkl} = \kappa_{sijkl'} \tag{2.14.4}$$

$$\lambda_{Sijkl} = \lambda_{Sijkl'}, \qquad k > 1 \tag{2.14.5}$$

$$\lambda_{CMij2l} = \lambda_{CMij2l'} \tag{2.14.6}$$

$$\lambda_{UMij2l} = \lambda_{UMij2l'} \tag{2.14.7}$$

$$\lambda_{Mijkl} = \lambda_{Mijkl'}, \qquad k > 2 \tag{2.14.8}$$

$$\alpha_{ijkl} = \alpha_{ijkl'} \tag{2.14.9}$$

and in case of common latent state variables S_{tjl}, additionally, $\delta_{ij1l} = 0$ and

$$\lambda_{Sij1l} = \lambda_{Sij1l'} \tag{2.14.10}$$

The definition of MI in Definition 2.4 corresponds to strict factorial invariance in the terms of Meredith (1993), as the residual variances of the LS-Com GRM are invariant by definition ($\theta_{ijkl} = 1$ for all i, j, k, and l). Definition 2.4 states that MI in the indicator-specific LS-Com GRM holds if all of the loading parameters for the latent factors S_{tij1l}, CM_{tj2l}, UM_{rtj2l} and M_{tjkl} as well as the threshold parameters κ_{sijkl} of all categories $s \in S_{ij}$ are invariant over measurement occasions for every indicator i of each construct j and method k. Note that this definition does include the loading parameters λ_{Sijkl}, $k > 1$, and the intercept parameters α_{ijkl}, that is, those parameters that represent the regression coefficients of the regression of the non-reference method indicators on the reference method indicators. Invariance of these parameters indicates that the conditional method bias is invariant over time, which would, strictly speaking, not be necessary to ensure that the latent variables S_{tij1l}, CM_{tj2l}, UM_{rtj2l} and M_{tjkl} have the same meaning over time, is, however, needed to compare differences in observed test scores over time.

Under strict factorial invariance, differences in observed test scores can be interpreted as differences in the underlying attributes (Meredith & Teresi, 2006). Also, under strict factorial invariance, differences in the means of observed test scores between measurement occasions are due to differences in the latent state means. That is, under strict factorial invariance it is legitimate to compare observed test scores between occasions of measurement. In contrast, observed differences in test scores might be confounded with differences in item-specific loadings or thresholds in cases where full MI is not given. If strict factorial invariance holds for specific items only, the above interpretations are valid for the respective items only.

While differences in observed responses can not necessarily be interpreted as differences in the underlying latent variables if MI is not given, differences in the latent state means per se might still be interpretable under certain conditions. For comparisons of the latent state means $\mathbb{E}(S_{tij1l})$ between different measurement occasions, the invariance settings as well as the choice of identification constraints have to be taken into account. As shown in Section 2.12, the expected value of the common latent response variables π_{tijkl} and π_{rtij2l} equal the expectation of the latent state factors S_{tij1l} if and only if $\lambda_{Sijkl} = 1$. For instance, for an item belonging to the reference method, equality of the latent state means $\mathbb{E}(S_{tij1l})$ over measurement occasions only corresponds to an equality in the probability curves $P(Y_{tij1l} \geq s \mid p_T, p_{TS_l}) = \Phi(\pi_{tsij1l}) = \Phi(\pi_{tij1l} - \kappa_{sij1l})$, for a given i and j, if the thresholds for the respective category s, κ_{sijkl}, as well as the loading parameters for the respective item λ_{Sij1l} are identical. That is, if one and only one threshold κ_{sijkl} for a chosen s is set invariant over different measurement occasions l for the reference item ($\lambda_{Sijkl} = 1$) of construct j, the difference in latent state means $\mathbb{E}(S_{tij1l}) - \mathbb{E}(S_{tij1l'})$ corresponds to the mean difference of the latent response variables π_{tsijkl} and $\pi_{tsijkl'}$ for the chosen category s, item i, and method k, and should only be interpreted as such. If, however, all threshold parameters κ_{sijkl} of the reference item (but not any other items) of construct j are constrained to be invariant over different measurement occasions, the difference in latent state means $\mathbb{E}(S_{tij1l})$ and $\mathbb{E}(S_{tij1l'})$ corresponds to the mean difference of the latent response variables π_{tsijkl} and $\pi_{tsijkl'}$ for all categories s of that item. That is, it can be interpreted as the mean difference of the latent state variable for the respective item i as measured by the respective method k.

Note that an unambiguous interpretation of other model parameters is already possible under weak factorial invariance for the latent response variables Y^*_{tijkl} and Y^*_{rtij2l}, as defined by Meredith (1993). That is, constant loading parameters for a latent factor with non-invariant threshold parameters allow for a meaningful interpretation of, e.g., the correlations of the respective factor between different measurement occasions. The violation of weak factorial invariance, however, would imply that the respective factor is not measured in the same way, rendering the interpretation of correlations difficult (Geiser, 2008). Note that for correlations between the latent state factors S_{j1l} of one construct j at different occasions to be meaningfully interpretable, only the loading parameters on the reference method have to be invariant, as these define the meaning of the latent state variable.

Chapter 3

Latent Change (LC-Com) Graded Response Model

Latent change (LC) models (McArdle & Hamagami, 2001; Steyer, Eid, & Schwenkmezger, 1997; Steyer, Partchev, & Shanahan, 2000) incorporate latent difference variables as change factors in longitudinal SEMs. The basic idea of LC models is that a latent state variable at a measurement occasion $l > 1$ can be decomposed into an initial state factor and a latent difference state factor:

$$S_{tij1l} = S_{tij11} + (S_{tij1l} - S_{tij11}), \quad l > 1 \tag{3.1.1}$$

with $S^{BC}_{tij1l} := (S_{tij1l} - S_{tij11})$. The latent difference state factor S^{BC}_{tij1l} represents the latent change between measurement occasions l and 1. Latent change models in which latent change is modeled with respect to the first measurement occasion are also called baseline change (BC) models (Steyer et al., 2000). Alternatively, latent difference variables could be defined to represent true change between adjacent measurement occasion (so-called neighbor change models; Steyer et al., 1997, 2000). The idea of defining latent difference factors can also be applied to the method factors of the LS-Com GRM (Geiser et al., 2010; Koch, 2013), e.g. for the structurally different non-reference method,

$$M_{tijkl} = M_{tijk1} + (M_{tijkl} - M_{tijk1}), \quad l > 1, k > 2 \tag{3.1.2}$$

with $M^{BC}_{tijkl} := (M_{tijkl} - M_{tijk1})$. The latent method difference variables M^{BC}_{tijkl} represent change in the non-reference methods that is not accounted for by change in the reference method (Geiser, 2008).
In contrast to latent state (i.e., multi-state) models, which can model inter-individual differences in intra-individual change only indirectly via the correlations of the state (and method) factors over time, latent change models allow to directly measure differential change. The latent difference variables indicate the variability in latent change between persons, that is, the latent change variables will have non-zero variances only if there are differences in change between individuals (Geiser, 2008). Multimethod latent change models, in addition, allow to study inter-individual differences in intra-individual change simultaneously for different methods and investigate the convergent validity in the assessment of change (Geiser et al., 2010; Koch, 2013).
LC models represent an alternative parameterization of latent state models, as can be seen from the tautological decompositions in Equations (3.1.1) and (3.1.2). As reformulations of latent state models they do not impose any additional assumptions. However, latent difference variables can only be meaningfully interpreted if factor loadings of the latent state and method variables as well as threshold parameters are time-invariant (Geiser, 2008), implying that the same construct is measured at the different measurement occasions. That is, strong measurement invariance has to hold in the latent state model (Koch, 2013; Steyer et al., 2000).

A latent change version of the continuous-indicator LS-Com model (LC-Com model) was defined in detail by Koch (2013). If strong measurement invariance holds, the LC-Com model is mathematically equivalent to the LS-Com model (see Koch, 2013). Consequently, this also holds for an LC version of the LS-Com GRM (see 2.4 for a definition of strong MI in the LS-Com GRM). That is, formal definitions of the latent change variables in the LC-Com GRM, their uniqueness, admissible transformations and meaningful statements, as well as the covariance structure of the model and variance decomposition are identical to the LC-Com model as defined by Koch (2013), however building on the definition of the latent state and method variables in the LS-Com GRM given in Chapter 2. Uniqueness, admissible transformations and meaningful statements for the remaining variables (e.g., thresholds and latent response variables) correspond to those in the LS-Com GRM (see Section 2.5). Furthermore, independence assumptions and identifiability conditions for the LC-Com GRM also correspond to those of the LS-Com GRM (see Sections 2.9 and 2.13).
An LC-Com GRM with indicator-specific latent state and change variables S_{tij1} and S^{BC}_{tijk} is depicted in Figure 3.1, an LC-Com GRM with common latent state and change variables S_{tj1} and S^{BC}_{tjk} is depicted in Figure 3.2.

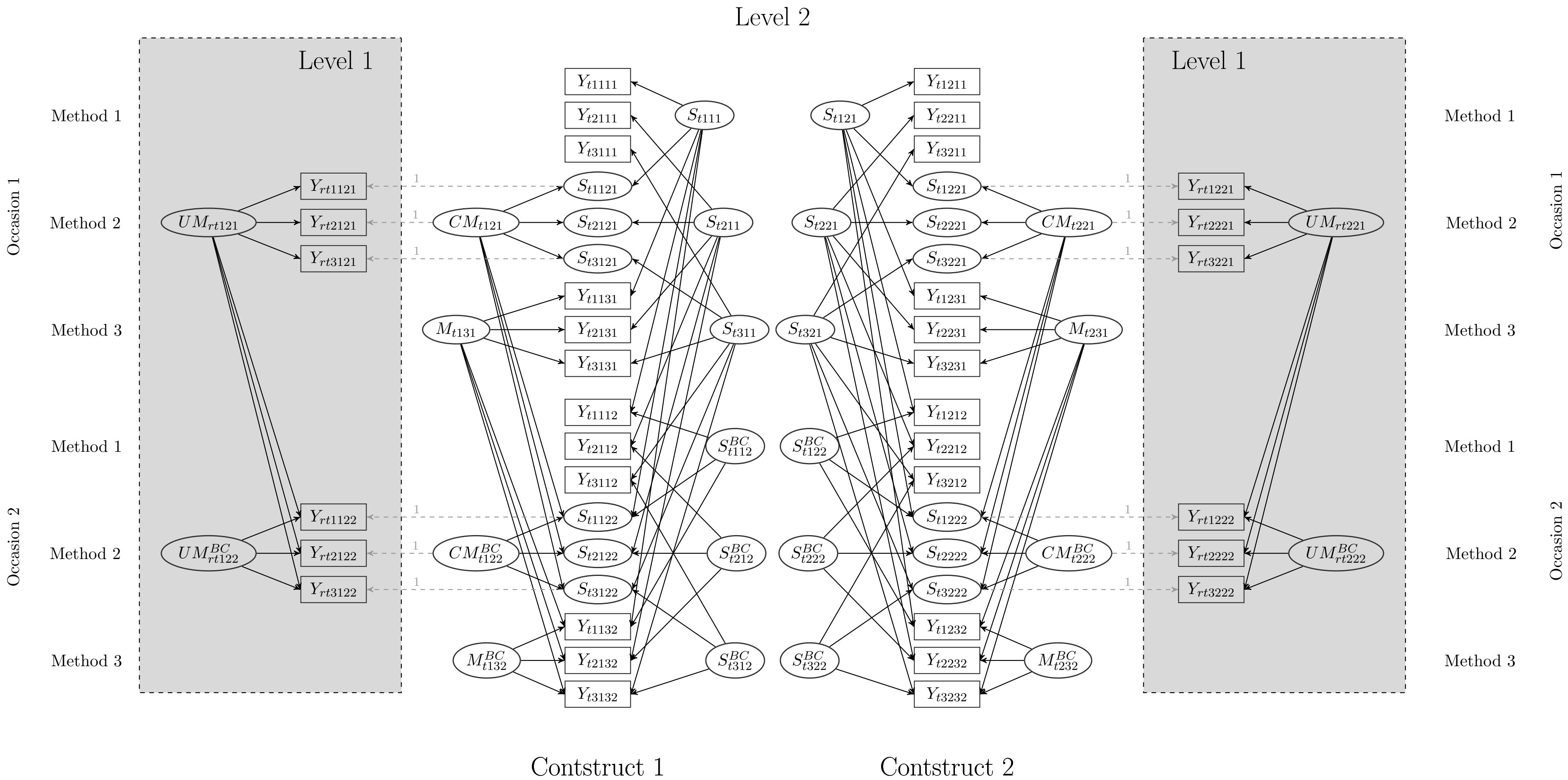

Figure 3.1: Path diagram of the Latent-Change-Com graded response model with indicator-specific latent state and change variables, S_{tij1} and S^{BC}_{tijk}. The model is depicted for two structurally different methods and one set of interchangeable methods at two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For the sake of clarity, correlations between latent variables and loading parameters are omitted. Note that loading parameters in the LC-Com GRM are time invariant, e.g., $\lambda_{Sijkl} = \lambda_{Sijkl'} \; \forall \; l, l'$, and analogously for the method factor loadings, due to measurement invariance restrictions. Furthermore, loading parameters of the latent (method) change factors correspond to those of the respective state / method factor for the same item i and construct j. Correlations that are not permissible in the LC-Com GRM are correlations between the latent state variables S_{tij1} and the latent (common) methods variables CM_{tj21} and M_{tj31} of the same construct j, as well as correlations between any level-1 and any level-2 latent variable. *BC*: baseline change variables; *CM*: common method variable; *M*: method variable; *S*: latent state variable; *UM*: unique method variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l.

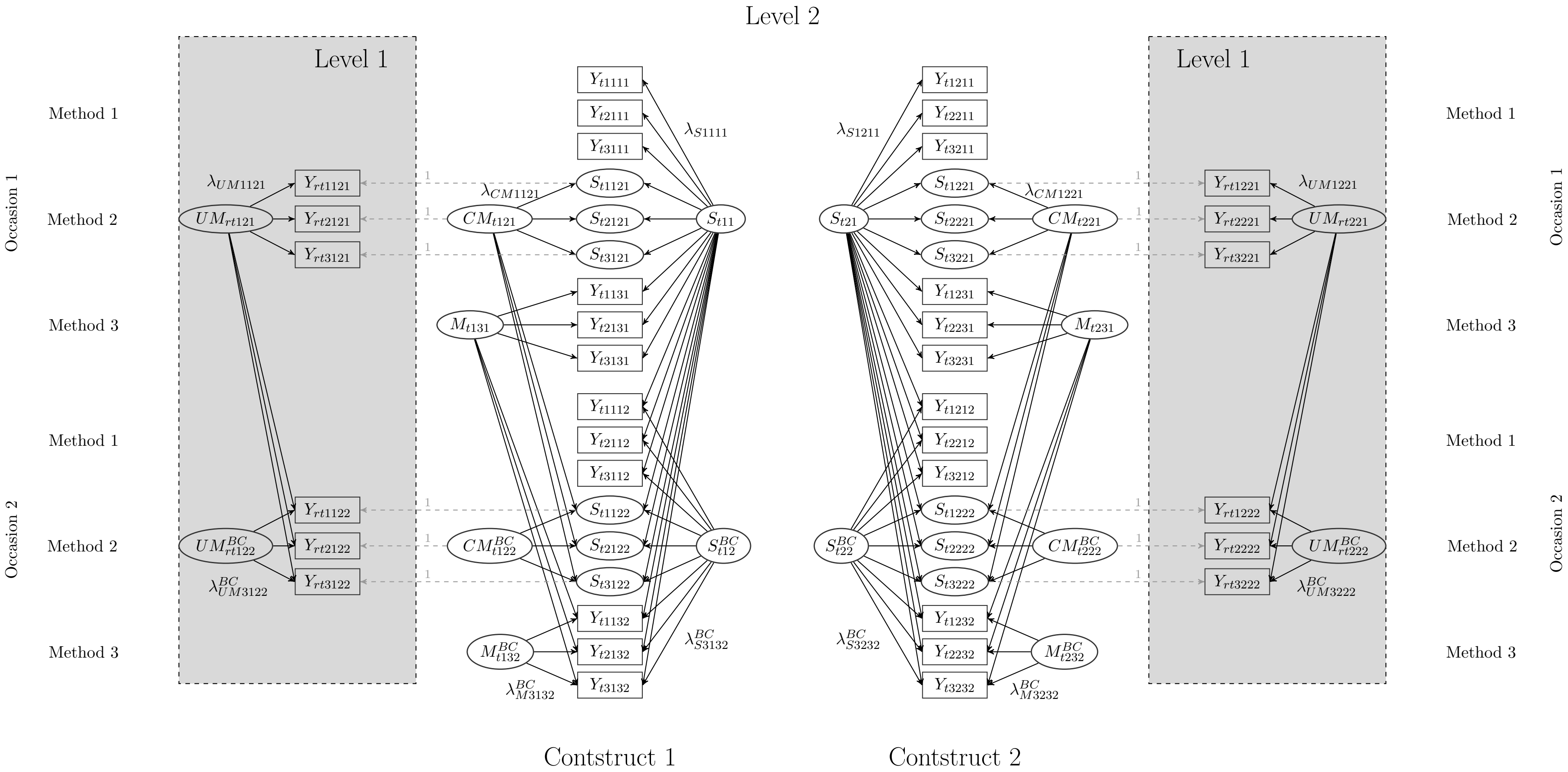

Figure 3.2: Path diagram of the Latent-Change-Com graded response model with common latent state and change variables, S_{tj1} and S^{BC}_{tjk}. The model is depicted for two structurally different methods and one set of interchangeable methods at two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For the sake of clarity, correlations between latent variables and loading parameters are omitted. Note that loading parameters in the LC-Com GRM are time invariant, e.g., $\lambda_{Sijkl} = \lambda_{Sijkl'} \; \forall \; l, l'$, and analogously for the method factor loadings, due to measurement invariance restrictions. Furthermore, loading parameters of the latent (method) change factors correspond to those of the respective state / method factor for the same item i and construct j. Correlations that are not permissible in the LC-Com GRM are correlations between the latent state variables S_{tj1} and the latent (common) methods variables CM_{tj21} and M_{tj31} of the same construct j, as well as correlations between any level-1 and any level-2 latent variable. BC: baseline change variables; CM: common method variable; M: method variable; S: latent state variable; UM: unique method variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l.

Chapter 4

Latent State-Trait (LST-Com) Graded Response Model

4.1 Introduction to the LST-Com GRM

This chapter introduces a longitudinal multilevel MTMM latent state-trait graded response model for measurement designs combining structurally different and interchangeable methods (LST-Com GRM). The model is based on the LST-Com model for continuous indicator variables developed by Koch et al. (2017). Furthermore, the definition of the latent response variables in the LST-Com GRM builds on the definition of the LS-Com GRM given in section 2.4. That is, the random experiment that characterizes the sampling procedure for longitudinal measurement designs of a graded response model with structurally different and interchangeable methods is identical in the LS-Com GRM and LST-Com GRM (see Sections 2.3 and 2.4). Based on the specified random experiment, the variables in the LST-Com GRM can then be properly defined as random variables.
LST models are widely applied in psychology and the social sciences (see Geiser & Lockhart, 2012, for an overview of LST applications). Despite many extensions of LST models proposed in the last decades (Cole et al., 2005; Eid, 1996; Eid & Hoffmann, 1998; Eid & Langeheine, 1999, 2003; Eid, Schneider, & Schwenkmezger, 1999; Geiser & Lockhart, 2012; Hamaker, Nesselroade, & Molenaar, 2007; Schermelleh-Engel, Keith, Moosbrugger, & Hodapp, 2004, among others) only few models for combining LST theory and MTMM analyses were introduced (Courvoisier et al., 2008; Koch et al., 2017; Scherpenzeel & Saris, 2007; Vautier, 2004). While the model by Scherpenzeel and Saris (2007) is limited to single indicator measurement designs and assumes uncorrelated method factors, the model by Courvoisier et al. (2008) is a multiple-indicator LST model based on the CTC(M-1) modeling approach (Eid, 2000; Eid et al., 2003). However, the model by Courvoisier et al. (2008) is limited to measurement designs including only structurally different methods. The LST-Com model developed by Koch et al. (2017) overcomes this limitation by extending the model to measurement designs combining structurally different and interchangeable methods. Furthermore, an extension of LST models to polytomous item responses was introduced for single-method, single-level models (Eid, 1995, 1996). However, an extension of LST MTMM modeling approaches to polytomous item responses is yet missing.
The LST-Com GRM combines the modeling possibilities of the continuous indicator LST-Com model (i.e., a model integrating LST theory and MTMM modeling of a combination of interchangeable and structurally different methods) with the advantages of an IRT approach to analyzing longitudinal MTMM data. The LST-Com GRM allows to

1. analyze whether a construct is more trait-like (stable) or state-like (occasion-specific),

2. analyze whether method effects are more trait-like (stable) or state-like (occasion-specific),
3. analyze convergent validity on the trait-level as well as,
4. analyze convergent validity on the occasion-specific level,
5. investigate whether particular interchangeable raters deviate from the common view of the interchangeable raters on the trait- or occasion-specific level,
6. investigate the generalizability of time-stable method effects across methods,
7. investigate the generalizability of occasion-specific method effects across methods,
8. do all the foregoing analyses on the item-level, and
9. compare item difficulties, item discrimination and reliability coefficients for the different items and over methods.

That is, the LST-Com GRM overcomes some of the limitations of previous longitudinal LST MTMM (continuous-indicator) models (Courvoisier et al., 2008; Koch et al., 2017) or single-method LST GRMs (Eid, 1996) by combining LST-MTMM models for interchangeable and structurally different methods with an IRT approach.

Recently, Steyer et al. (2015) proposed a revision of LST theory (LST-R) that explicitly takes into account that persons might change over the course of time. Implications of LST-R theory for the present model are discussed in Section 4.13. Koch et al. (2017) show which modifications have to be made in order to make the (continuous-indicator) LST-Com model compatible with the revised version of LST theory.

The basic concept of LST theory (Steyer, Majcen, Schwenkmezger, & Buchner, 1989; Steyer & Schmitt, 1990; Steyer et al., 1992, 1999) consists of the decomposition of latent state variables into latent trait and latent state residuals. While the latent trait variables represent the person-specific expectation of the latent state variables over situations and measurement occasions, the latent state residual variables represent situation and / or person-situation interaction effects (Steyer et al., 1992, 1999). That is, LST theory provides the methodological framework for disentangling stable person-specific effects, time-variable effects and measurement error influences (Steyer et al., 1999).

To illustrate the concept, consider the reference-method, self-report latent response variables π_{tij1l}. Recall that these variables were defined as

$$\pi_{tij1l} = \pi_{tsij1l} + \kappa_{sij1l}$$

with

$$\pi_{tsij1l} := \Phi^{-1}[P(Y_{tij1l} \geq s \mid p_T, p_{TS_l})]$$

and

$$P(Y_{tij1l} \geq s \mid p_T, p_{TS_l}) = \mathbb{E}[I_{\{Y_{tij1l} \geq s\}} \mid p_T, p_{TS_l}]$$

(see Section 2.4). That is, the variables π_{tij1l} are (p_T, p_{TS_l})-measurable functions, as they can be defined as $\pi_{tij1l} : \varphi_{tij1l}(p_T, p_{TS_l})$, i.e., a composition of the mappings $(p_T, p_{TS_l}) : \Omega \to \Omega_T \times \Omega_{TS_l}$ and $\varphi_{tij1l} : \Omega_T \times \Omega_{TS_l} \to \mathbb{R}$. It follows that the latent state variable $S_{tij1l} = \pi_{tij1l}$ is a (p_T, p_{TS_l})-measurable function. The latent variables in an LST GRM are then defined as follows (cf. Eid, 1996; Koch et al., 2017):

$$\xi_{tij1l} = \mathbb{E}[S_{tij1l} \mid p_T] \tag{4.1.1}$$

$$\zeta_{tij1l} = S_{tij1l} - \xi_{tij1l} \tag{4.1.2}$$

That is, the latent state variable S_{tij1l} is decomposed into a latent trait variable ξ_{tij1l} and a latent state residual variable ζ_{tij1l}:

$$S_{tij1l} = \xi_{tij1l} + \zeta_{tij1l} \tag{4.1.3}$$

The latent trait variables are defined as the conditional expectations of the latent state variables given the (target) person projection p_T. The latent trait variable characterizes the person itself, that is, it represents the value of the latent state variable one would expect for a person t on measurement occasion l, irrespective of the specific situation realized on that measurement occasion (Eid, 1996). The latent state residual variables ζ_{tij1l} are latent occasion-specific variables that are defined as latent residuals with respect to the trait variables ξ_{tij1l}. They are the occasion-specific deviations of the latent variable S_{tij1l} from the (target-) person-specific expectations of that variable and thereby represent influences of the situation and person-situation interactions.
In the LST-Com GRM, this decomposition into trait and state residual components is extended to the non-reference method indicators, allowing for the definition of time-stable and occasion-specific method effects. On Level 2, the latent trait variable of a non-reference method (ξ_{tijkl}, $k \neq 1$) is regressed on the latent trait variable of the reference method (ξ_{tij1l}) of the same indicator i of construct j at occasion l, according to the CTC(M-1) approach (Courvoisier et al., 2008; Eid, 2000; Koch, 2013). The residuals of these regressions are defined as latent trait (common) method variables:

$$\xi^{CM}_{tij2l} = \xi_{tij2l} - \mathbb{E}[\xi_{tij2l} \mid \xi_{tij1l}] \tag{4.1.4}$$

$$\xi^{M}_{tijkl} = \xi_{tijkl} - \mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}] \qquad \forall \ k > 2 \tag{4.1.5}$$

They reflect the stable, time-consistent view of the informant ratings that cannot be explained by the stable components of the reference method reports (e.g., self-reports). Analogously, the latent state residual variables of the non-reference methods (ζ_{tijkl}, $k \neq 1$) can be regressed on the latent state residual variable of the reference method (ζ_{tij1l}) of the same indicator i of construct j at occasion l to yield latent state residual (common) method variables:

$$\zeta^{CM}_{tij2l} = \zeta_{tij2l} - \mathbb{E}[\zeta_{tij2l} \mid \zeta_{tij1l}] \tag{4.1.6}$$

$$\zeta^{M}_{tijkl} = \zeta_{tijkl} - \mathbb{E}[\zeta_{tijkl} \mid \zeta_{tij1l}] \qquad \forall \ k > 2 \tag{4.1.7}$$

They represent the part of the momentary, occasion-specific view of the informant method that cannot be explained by the occasion-specific view of the reference method on the same measurement occasion (Koch et al., 2017). Note that the regressions of the non-reference method trait / state residual variables on the reference-method trait / state residual variables are assumed to be linear.
Furthermore, the level-1 residuals UM_{rtij2l} can be decomposed into a time-consistent expectation given the person-projection p_T and the rater-projection p_R and an occasion-specific residual variable representing situational effects and rater-situation interactions:

$$\xi^{UM}_{rtij2l} = \mathbb{E}[UM_{rtij2l} \mid p_T, p_R] \tag{4.1.8}$$

$$\zeta^{UM}_{rtij2l} = UM_{rtij2l} - \xi^{UM}_{rtij2l} \tag{4.1.9}$$

Formal definitions and detailed explanations of the latent variables and assumptions made in the LST-Com GRM are presented in the following sections. An LST-Com GRM with indicator-specific latent trait (method) and state residual factors is depicted in Figure 4.1, an LST-Com GRM with common latent trait (method) and state residual factors in Figure 4.2.
The decomposition of the latent response variables (as defined in Section 2.4) in the LST-Com GRM, then lead to the following measurement equations (Koch et al., 2017):

$$\pi_{tij1l} = \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\xi_{tij1} + \zeta_{tij1l} \tag{4.1.10}$$

$$\pi_{rtij2l} = \alpha_{\xi ij2l} + \lambda_{\xi ij2l}\xi_{tij1} + \lambda^{CM}_{\xi ij2l}\xi^{CM}_{tij2} + \lambda_{\zeta ij2l}\zeta_{tij1l} + \lambda^{CM}_{\zeta ij2l}\zeta^{CM}_{tj2l} + \lambda^{UM}_{\xi ij2l}\xi^{UM}_{rtij2} + \lambda^{UM}_{\zeta ij2l}\zeta^{UM}_{rtj2l} \tag{4.1.11}$$

$$\pi_{tijkl} = \alpha_{\xi ijkl} + \lambda_{\xi ijkl}\xi_{tij1} + \lambda^{M}_{\xi ijkl}\xi^{M}_{tijk} + \lambda_{\zeta ijkl}\zeta_{tij1l} + \lambda^{M}_{\zeta ijkl}\zeta^{M}_{tjkl} \qquad k > 2 \tag{4.1.12}$$

Here, the variables ξ_{tij1}, ξ^{CM}_{tij2}, ξ^{M}_{tijk}, and ξ^{UM}_{rtij2} are assumed to be time-stable latent trait (method) variables and hence do not have a time index l any more. As shown in the following sections, the latent state residual variables as well as all method variables are defined as latent residuals and are thereby uncorrelated with their regressors. This allows the definition of different variance components of the latent response variables. The time consistency and occasion-specificity coefficients are analogous to the consistency and specificity coefficients specified in classical LST theory models (Steyer & Schmitt, 1990; Steyer et al., 1992), with the exception that they are latent variance components (i.e., defined for the latent response variable, not including measurement error). The coefficients indicating the convergence between different methods (consistency and method specificity) are defined analogously to the coefficients defined for multilevel CFA and CTC(M-1) models of MTMM data in Eid et al. (2008), with the possibility to define method consistency and specificity on both the trait and the occasion-specific levels in the LST-Com model. Most of these coefficients are defined in the same way in Koch et al. (2017). The time consistency coefficients, given for the reference-method and interchangeable method, $k = 2$, here,

$$Con(\pi_{tij1l}) = \frac{(\lambda_{\xi ij1l})^2 Var(\xi_{tij1})}{Var(\pi_{tij1l})} \tag{4.1.13}$$

$$Con(\pi_{rtij2l}) = \frac{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2}) + (\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})}{Var(\pi_{rtij2l})} \tag{4.1.14}$$

represent the proportion of variance of the latent response variable that goes back to stable (not occasion-specific or momentary) influences. In contrast, the occasion specificity coefficient indicates how much of the variance in the latent response variables is attributable to momentary, occasion-specific influences:

$$OSpe(\pi_{tij1l}) = \frac{Var(\zeta_{tij1l})}{Var(\pi_{tij1l})} \tag{4.1.15}$$

$$OSpe(\pi_{rtij2l}) = \frac{(\lambda_{\zeta ij2l})^2 Var(\zeta_{tij1l}) + (\lambda^{CM}_{\zeta ij2l})^2 Var(\zeta^{CM}_{tj2l}) + (\lambda^{UM}_{\zeta ij2l})^2 Var(\zeta^{UM}_{rtj2l})}{Var(\pi_{rtij2l})} \tag{4.1.16}$$

The trait method consistencies are indicators of the convergent validity on the stable trait-level, that is, they indicate how much of the stable variance in the non-reference methods can be explained by stable inter-individual differences in the reference-method, e.g., for the interchangeable non-reference method:

$$TCon(\pi_{rtij2l}) = \frac{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1})}{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2}) + (\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})} \tag{4.1.17}$$

$$TCon(\pi_{tij2l}) = \frac{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1})}{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2})} \tag{4.1.18}$$

While $TCon(\pi_{rtij2l})$ is the proportion of stable variance in the individual peer reports that is attributable to stable interindividual differences measured by the reference method, $TCon(\pi_{tij2l})$ represents the proportion of stable variance in the common view of the interchangeable peer reports that is attributable to stable interindividual differences measured by the reference method. The square root of the trait (method) consistency coefficients can be interpreted as the degree of true convergent validity on the trait-level (Koch et al., 2017). The trait method specificity coefficients represent the proportion of variance that is determined by stable method-specific influences of the non-reference methods.

$$TUMS(\pi_{rtij2l}) = \frac{(\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})}{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2}) + (\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})} \quad (4.1.19)$$

$$TCMS(\pi_{rtij2l}) = \frac{(\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2})}{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2}) + (\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})} \quad (4.1.20)$$

$$TCMS(\pi_{tij2l}) = \frac{(\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2})}{(\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2})} \quad (4.1.21)$$

$$TMS(\pi_{tijkl}) = \frac{(\lambda^{M}_{\xi ijkl})^2 Var(\xi^{M}_{tijk})}{(\lambda_{\xi ijkl})^2 Var(\xi_{tij1}) + (\lambda^{M}_{\xi ijkl})^2 Var(\xi^{M}_{tijk})} \quad (4.1.22)$$

The trait unique method specificity $TUMS(\pi_{rtij2l})$ represents the stable view of a peer, which cannot be explained by stable inter-individual differences in the reference method (e.g., self-reports) and is not shared with the other interchangeable raters (Koch et al., 2017). The trait common method specificity $TCMS(\pi_{rtij2l})$, in contrast, represents the proportion of stable variance in the individual peer reports that cannot be explained by stable inter-individual difference in the reference method but is shared with the other interchangeable raters. The trait common method specificity $TCMS(\pi_{tij2l})$ is the proportion of the stable common view of all interchangeable peers that is not shared with the stable view of the reference-method raters. All of the preceding method consistency and specificity coefficients can also be computed on the occasion-specific, momentary level. E.g.,

$$OCon(\pi_{tijkl}) = \frac{(\lambda_{\zeta ijkl})^2 Var(\zeta_{tij1l})}{(\lambda_{\zeta ijkl})^2 Var(\zeta_{tij1l}) + (\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl})} \quad (4.1.23)$$

$$OMS(\pi_{tijkl}) = \frac{(\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl})}{(\lambda_{\zeta ijkl})^2 Var(\zeta_{tij1l}) + (\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl})} \quad (4.1.24)$$

define the occasion-specific method consistency and occasion-specific method specificity for the structurally different non-reference method. $OCon(\pi_{tijkl})$ represents the part of the momentary, occasion-specific variance of a non-reference method indicator that can be explained by the momentary, occasion-specific inter-individual differences in the reference-method reports. The square root of the occasion-specific (method) consistency coefficient can be interpreted as the degree of true convergent validity on the occasion-specific level (Koch et al., 2017). $OMS(\pi_{tijkl})$ represents the occasion-specific variance that is attributable to the momentary view of the structurally different raters but not shared with the reference-method report (e.g., self-report).
For definitions of these coefficients for the remaining methods see Section 4.9. For a more detailed explanation of the variance coefficients' meaning see Koch et al. (2017) or Koch (2013).
The latent variables in the LST-Com GRM and assumptions underlying the model definitions are formally defined in the following sections. The uniqueness of the latent variables and their coefficients, admissible transformations and meaningful statements regarding the former are discussed, necessary independence assumptions are introduced and testable consequences for the covariance structure of the model are derived. Last but not least, identification conditions for the model are presented and conditions for testing measurement invariance in the LST-Com GRM are discussed.

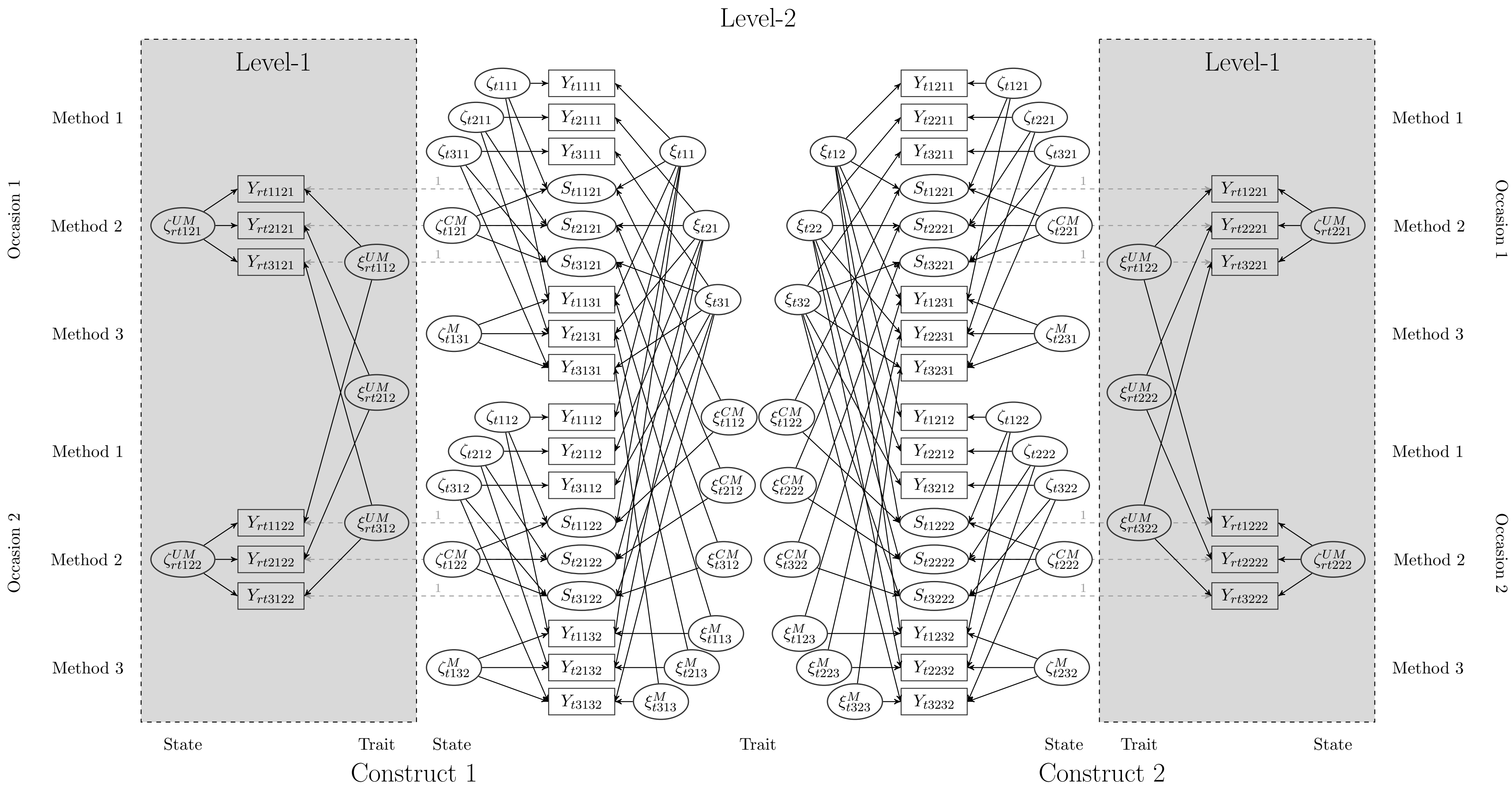

Figure 4.1: Path diagram of the LST-Com graded response model with indicator-specific latent trait variables ξ_{tij}, ξ^{UM}_{rtij2}, ξ^{CM}_{tij2} and ξ^{M}_{tijk} and latent state residual variables ζ_{tijl}. The model is depicted for two structurally different methods and one set of interchangeable methods on two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For convenience, the constant indicator $k = 1$ has been dropped from the latent trait variables (ξ_{tij}=ξ_{tij1}) and the latent state residual variables (ζ_{tijl}=ζ_{tij1l}). For the sake of clarity, correlations between latent variables and loading parameters are omitted. Correlations that are not permissible in the depicted LST-Com GRM are all correlations between any trait (method) variable ξ and any state residual (method) variable ζ, correlations between the latent trait and the latent trait (common) method variables of the same construct j and indicator i, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. *CM*: common method; *M*: method; *S*: state variable; *UM*: unique method; ξ: latent trait variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

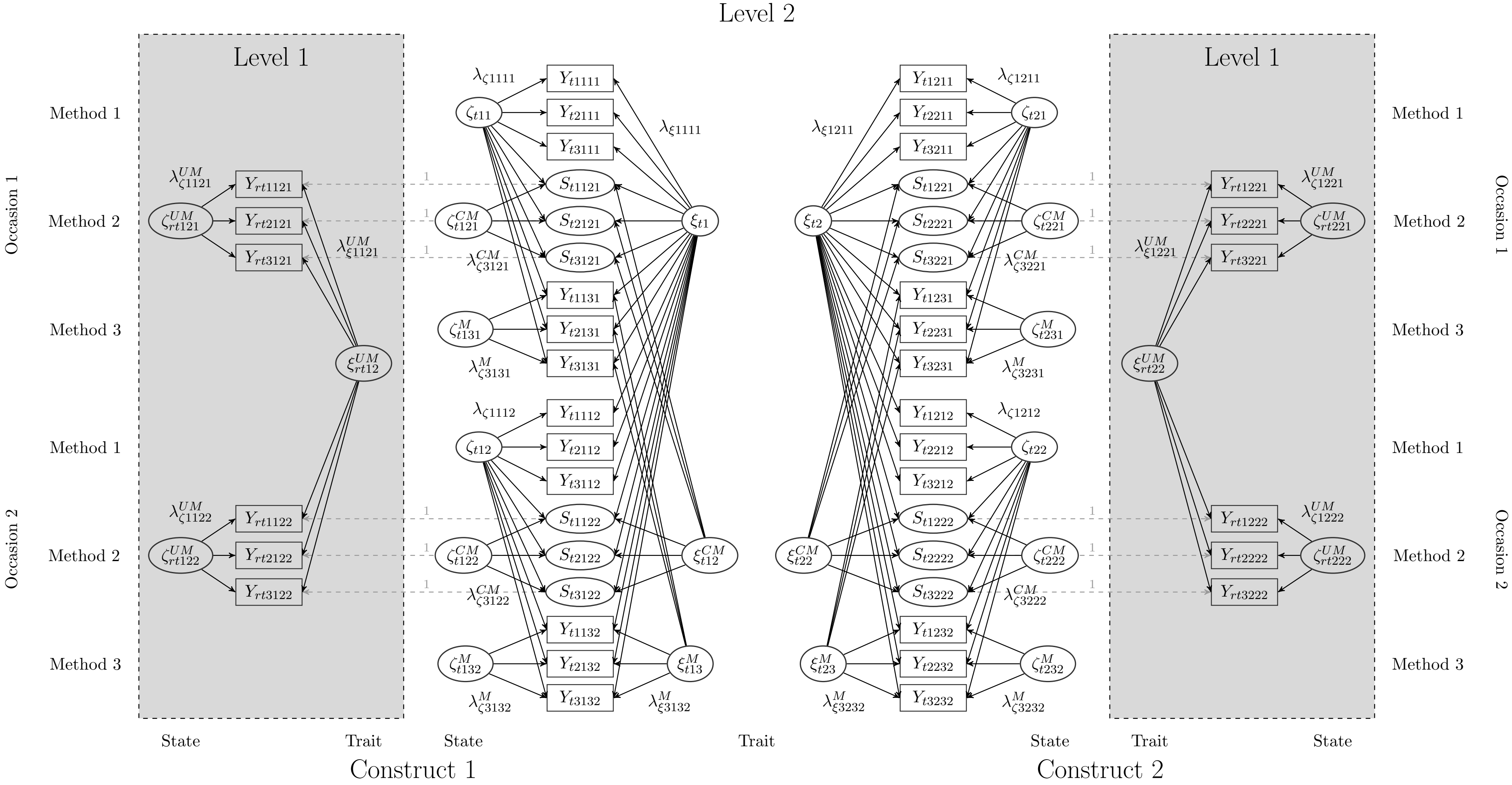

Figure 4.2: Path diagram of the LST-Com graded response model with common latent trait variables ξ_{tj}, ξ^{UM}_{rtj2}, ξ^{CM}_{tj2} and ξ^{M}_{tjk} and latent state residual variables ζ_{tjl}. The model is depicted for two structurally different methods and one set of interchangeable methods on two measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For convenience, the constant indicator $k = 1$ has been dropped from the latent trait variables (ξ_{tj}=ξ_{tj1}) and the latent state residual variables (ζ_{tjl}=ζ_{tj1l}). For the sake of clarity, correlations between latent variables are omitted and loading parameters are only shown for exemplary indicators. Correlations that are not permissible in the depicted LST-Com GRM are all correlations between any trait (method) variable ξ and any state residual (method) variable ζ, correlations between the latent trait and the latent trait (common) method variables of the same construct j, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method; M: method; S: state variable; UM: unique method; ξ: latent trait variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

4.2 Formal Definition of the LST-Com GRM

In the following the LST-Com GRM is formally defined building on the definition of the LS-Com GRM in section 2.4 and the LST-Com model for continuous indicators (Koch, 2013; Koch et al., 2017).

Definition 4.1. (LST-Com GRM)

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ *and* $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ *on a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *are variables of an LST-Com graded response model if conditions* (a) *to* (e) *in Definition 2.1 [i.e., LS-Com GRM] and the following conditions hold:*

(a) The following variables are random variables on $(\Omega, \mathcal{A}, \mathcal{P})$ *with finite first- and second-order moments:*

Rater-level (Level 1):

$$UM_{rtij2l} = \pi_{rtij2l} - \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \tag{4.2.1}$$

$$\xi^{UM}_{rtij2l} = \mathbb{E}[UM_{rtij2l} \mid p_T, p_R] \tag{4.2.2}$$

$$\zeta^{UM}_{rtij2l} = UM_{rtij2l} - \xi^{UM}_{rtij2l} \tag{4.2.3}$$

Target-level (Level 2):

$$S_{tijkl} = \pi_{tijkl} \qquad \forall\ k \neq 2, \tag{4.2.4}$$

$$S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}] \tag{4.2.5}$$

$$\xi_{tijkl} = \mathbb{E}[S_{tijkl} \mid p_T] \tag{4.2.6}$$

$$\zeta_{tijkl} = S_{tijkl} - \xi_{tijkl} \tag{4.2.7}$$

$$\xi^{CM}_{tij2l} = \xi_{tij2l} - \mathbb{E}[\xi_{tij2l} \mid \xi_{tij1l}] \tag{4.2.8}$$

$$\xi^{M}_{tijkl} = \xi_{tijkl} - \mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}] \qquad \forall\ k > 2 \tag{4.2.9}$$

$$\zeta^{CM}_{tij2l} = \zeta_{tij2l} - \mathbb{E}[\zeta_{tij2l} \mid \zeta_{tij1l}] \tag{4.2.10}$$

$$\zeta^{M}_{tijkl} = \zeta_{tijkl} - \mathbb{E}[\zeta_{tijkl} \mid \zeta_{tij1l}] \qquad \forall\ k > 2 \tag{4.2.11}$$

(b) For each indicator i of construct j on measurement occasion l, measured by a non-reference method $(k \neq 1)$*, there are constants* $\alpha_{ijkl} \in \mathbb{R}$ *and* $\lambda_{ijkl} \in \mathbb{R}^+$ *such that*

$$\mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}] = \alpha_{ijkl} + \lambda_{ijkl}\xi_{tij1l} \tag{4.2.12}$$

(c) *For each indicator i of construct j, measured by a non-reference method* $(k \neq 1)$ *on occasion of measurement l, there is a constant* $\lambda_{\zeta ijkl} \in \mathbb{R}^+$ *such that*

$$\mathbb{E}[\zeta_{tijkl} \mid \zeta_{tij1l}] = \lambda_{\zeta ijkl}\zeta_{tij1l} \tag{4.2.13}$$

(d) *Definition of common trait variables. For each indicator i of construct j, measured by the reference method* $(k = 1)$ *and for each pair* $(l,l') \in L \times L$, $l \neq l'$, *there are constants* $\alpha_{ij1ll'} \in \mathbb{R}$ *and* $\lambda_{ij1ll'} \in \mathbb{R}^+$ *such that*

$$\xi_{tij1l} = \alpha_{ij1ll'} + \lambda_{ij1ll'}\xi_{tij1l'} \tag{4.2.14}$$

(e) *Definition of common method trait variables. For each indicator i of construct j, measured by a non-reference method k* $(k \neq 1)$, *and for each pair* $(l,l') \in L \times L$, $(l \neq l')$, *there are constants* $\lambda^{UM}_{\xi ij2ll'}$, $\lambda^{CM}_{\xi ij2ll'}$ *and* $\lambda^{M}_{\xi ijkll'} \in \mathbb{R}^+$ *such that*

$$\xi^{UM}_{rtij2l} = \lambda^{UM}_{\xi ij2ll'}\xi^{UM}_{rtij2l'} \tag{4.2.15}$$

$$\xi^{CM}_{tij2l} = \lambda^{CM}_{\xi ij2ll'}\xi^{CM}_{tij2l'} \tag{4.2.16}$$

$$\xi^{M}_{tijkl} = \lambda^{M}_{\xi ij2ll'}\xi^{M}_{tijkl'} \qquad \forall\ k > 2 \tag{4.2.17}$$

(f) *Definition of common method state residual variables. For each construct j, measured by a non-reference method k* $(k \neq 1)$, *and for each pair* $(i,i') \in I_j \times I_j$, $(i \neq i')$, *there are constants* $\lambda^{UM}_{\zeta ii'j2l}$, $\lambda^{CM}_{\zeta ii'j2l}$ *and* $\lambda^{M}_{\zeta ii'jkl} \in \mathbb{R}^+$ *such that*

$$\zeta^{UM}_{rtij2l} = \lambda^{UM}_{\zeta ii'j2l}\zeta^{UM}_{rti'j2l} \tag{4.2.18}$$

$$\zeta^{CM}_{tij2l} = \lambda^{CM}_{\zeta ii'j2l}\zeta^{CM}_{ti'j2l} \tag{4.2.19}$$

$$\zeta^{M}_{tijkl} = \lambda^{M}_{\zeta ii'jkl}\zeta^{M}_{ti'jkl} \qquad \forall\ k > 2 \tag{4.2.20}$$

Remarks.
Definition 4.1 defines latent trait variables ξ_{tijkl}, latent state residual variables ζ_{tijkl}, as well as latent trait method variables ξ^{UM}_{rtij2l}, ξ^{CM}_{tij2l}, ξ^{M}_{tijkl} and latent state residual method variables ζ^{UM}_{rtij2l}, ζ^{CM}_{tij2l}, ζ^{M}_{tijkl}. Latent trait variables ξ_{tijkl} in the LST-Com GRM are defined as conditional expectations of the latent state variables S_{tijkl} given the target, i.e., $\mathbb{E}[S_{tijkl} \mid p_T]$. The latent trait variable characterizes the person itself, that is, it represents the value of the latent state variable one would expect for a person t on measurement occasion l, irrespective of the specific situation realized on measurement occasion l (Eid, 1996). Although trait scores are, by definition, not dependent on the specific situation realized on a measurement occasion, traits may nevertheless be subject to change over time (Steyer et al., 1999). That is, they are dependent on the probability distribution of the variables S_{tijkl}, which again depend on the probability distribution of the situations on measurement occasion l. The expectation of the probability distribution of the variables Y_{tijkl} and S_{tijkl} may change over time, as the probability distribution of the situations for a target t (or rater r) may have changed over time. Hence, the subscript l is not omitted in the definition of the latent trait and latent trait method variables ξ_{tijkl}, ξ^{UM}_{rtij2l}, ξ^{CM}_{tij2l} and ξ^{M}_{tijkl} in Equations (4.2.2), (4.2.6), (4.2.8) and (4.2.9). Steyer et al. (2015) have proposed a revised version of LST-theory (LST-R) that defines trait change on the basis of a different conceptualization of the random experiment. In contrast to LST-R theory (Steyer et al., 2015), the

current model definition assumes that it is possible to incorporate trait change into the model under this definition of the random experiment. The incorporation of occasion-specific trait variables into the current model definition is based on an understanding of situations and distributional assumptions for the situations and projections p_{S_l} that differ from those in LST-R theory. LST-R theory as proposed by Steyer et al. (2015) is discussed in section 4.13.

The latent state residual variables ζ_{tijkl} are latent occasion-specific variables that are defined as latent residuals of the latent state variables S_{tijkl} with respect to the trait variables ξ_{tijkl} (see Equation 4.2.7). That is, they are the occasion-specific deviations of the latent variable S_{tijkl} from the target-specific (and rater-specific) expectations of that variable on occasion l.

On Level 1, latent unique method trait variables ξ^{UM}_{rtij2l} are defined as target- and rater-specific expectations of the unique method variables, i.e., $\mathbb{E}[UM_{rtij2l} \mid p_T, p_R]$, while the latent unique method state residual variables ζ^{UM}_{rtij2l} are defined as residual variables of UM_{rtij2l} with respect to this expectation.

On Level 2, latent trait (common) method variables are defined as residual variables of the regression of the latent trait variable of a non-reference method (ξ_{tijkl}, $k \neq 1$) on the latent trait variable of the reference method (ξ_{tij1l}) of the same indicator i of construct j on occasion l (see Equations 4.2.8 and 4.2.9). Latent state residual (common) method variables, in contrast, are defined as the residuals of the latent regression of the latent state residual variable of the non-reference method (ζ_{tijkl}, $k \neq 1$) on the latent state residual variable of the reference method (ζ_{tij1l}) of the same indicator i of construct j on occasion l (see Equations 4.2.10 and 4.2.11). That is, the latent trait (common) method variables reflect the stable, time-consistent view of the informant ratings that cannot be explained by the stable components of the reference method reports (e.g., self-reports). The latent state residual (common) method variables, in contrast, are the part of the momentary, occasion-specific view of the informant method that cannot be explained by the occasion-specific view of the reference method on the same measurement occasion. Hence, it is possible to differentiate between momentary rater bias (as deviations in the occasion-specific views of the different raters) and stable, time-consistent method biases.

Note that the latent state residual variables ζ_{tijkl} as well as all of the latent trait method variables (ξ^{UM}_{rtij2l}, ξ^{CM}_{tij2l}, ξ^{M}_{tijkl}) and latent state residual method variables (ζ^{UM}_{rtij2l}, ζ^{CM}_{tij2l}, ζ^{M}_{tijkl}) are defined as latent residual variables and therefore have expectations of zero by definition. Hence, no additive constants are included in any of the equations in (c), (e) and (f) of definition 4.1 for these variables. The latent trait variables ξ_{tijkl}, in contrast, do not have zero expectations by definition, and Equations (4.2.12) and (4.2.14) do include intercept parameters. These intercept parameters and all of the coefficients κ_{sijkl} for the same i, j, k, and l are, however, not separately identifiable (see Section 4.3 for theorems on the uniqueness of the latent trait and latent response variables, as well as Section 4.11 on identifiability conditions).

Equation (4.2.12) states the assumption that the dependence of the non-reference method trait variables ξ_{tijkl}, $k \neq 1$, on the reference method trait variable ξ_{tij1l} of the same indicator i, construct j and occasion l can be described by linear transformations. Equation (4.2.13) states the assumption that the dependence of the non-reference method state residual variables ζ_{tijkl}, $k \neq 1$, on the reference method state residual variable ζ_{tij1l} of the same indicator i, construct j and occasion l can be described by similarity transformations.

Equation (4.2.14) formalizes the assumption that the reference method latent trait variables of different measurement occasions are linear transformations of each other. That is, they are assumed to be perfectly correlated. Hence, trait change that can be described by this model follows the same linear relationship for every target t (non-linear trait change or individual change trajectories for the different targets, i.e., inter-individual differences in intra-individual change, can be incorporated, yielding different models, such as the latent growth curve model, see chapter 5). Note that it is necessary to make an assumption of some form of dependence (linear or non-linear) between the latent trait (method) variables of different measurement occasions $l \neq l'$ $(l, l' \in L)$, as occasion-specific trait (method) variables and state residual (method) variables are not separately identifiable. That is, without the assumption of this dependence (as made for instance in Equation 4.2.14) it would not be possible to separate (occasion-specific) trait from occasion-specific state residual variance. Additionally, trait change may not be separable from some forms of measurement non-invariance (see section 4.4). However, see Eid and Hoffmann (1998) for an example on how to restructure the model to yield an unambiguous separation of trait change and measurement invariance.

Equations (4.2.15) - (4.2.17) define the assumptions that all latent trait method variables ξ^{UM}_{rtij2l}, ξ^{CM}_{tij2l}, and ξ^{M}_{tijkl}, $k > 2$, belonging to different measurement occasions $l \neq l'$ but the same indicator, construct and method are similarity transformations of each other, respectively. This assumption implies that the variables are perfectly correlated and can therefore be represented by common trait method factors (Koch, 2013). The existence of these common trait method factors is stated in Theorem 4.1 (cf. Koch, 2013).

Equations (4.2.18) - (4.2.20) define the assumptions that all latent state residual method variables ζ^{UM}_{rtij2l}, ζ^{CM}_{tij2l}, and ζ^{M}_{tijkl}, $k > 2$, belonging to different indicators $i \neq i'$ but the same construct, method and measurement occasion are similarity transformations of each other, respectively. This assumption implies that the variables are perfectly correlated and can therefore be represented by common state residual method factors ζ^{UM}_{rtj2l}, ζ^{CM}_{tj2l}, and ζ^{M}_{tjkl} per construct j and measurement occasion l (Koch, 2013). The existence of these common state residual method factors is stated in Theorem 4.1 (cf. Koch, 2013). Note that assumptions (4.2.18) and (4.2.20) are necessary for identifiability reasons, as a model with indicator-specific factors ζ^{UM}_{rtij2l} or ζ^{M}_{tijkl} would not be identified. This is not the case for the ζ^{CM}_{tij2l} variables, hence it is not necessary to impose assumption (4.2.19) to identify the model.

Theorem 4.1. ***(Existence)***

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ *and* $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ *are* $(\xi_{tij1l}, \xi^{CM}_{tij2l}, \xi^{UM}_{rtij2l}, \xi^{M}_{tijkl}, \zeta^{CM}_{tij2l}, \zeta^{UM}_{tij2l}, \zeta^{M}_{tijkl})$*-congeneric variables of an LST-Com GRM if and only if the conditions in Definition 4.1 hold. Then, for each* $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$, *there are real-valued random variables* ξ_{tij1}, ξ^{CM}_{tij2}, ξ^{UM}_{rtij2}, ξ^{M}_{tijk}, ζ^{CM}_{tj2l}, ζ^{UM}_{tj2l}, *and* ζ^{M}_{tjkl} *on a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *and constants* $\alpha_{\xi ijkl} \in \mathbb{R}$ *and* $(\lambda_{\xi ijkl}, \lambda^{CM}_{\xi ij2l}, \lambda^{UM}_{\xi ij2l}, \lambda^{M}_{\xi ijkl}, \lambda^{CM}_{\zeta ij2l}, \lambda^{UM}_{\zeta ij2l}, \lambda^{M}_{\zeta ijkl})$ $\in \mathbb{R}^+$ *such that:*

$$\xi_{tij1l} = \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\xi_{tij1} \tag{4.2.21}$$

$$\mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}] = \alpha_{\xi ijkl} + \lambda_{\xi ijkl}\xi_{tij1} \qquad k > 1 \tag{4.2.22}$$

$$\xi^{UM}_{rtij2l} = \lambda^{UM}_{\xi ij2l}\xi^{UM}_{rtij2} \tag{4.2.23}$$

$$\xi^{CM}_{tij2l} = \lambda^{CM}_{\xi ij2l}\xi^{CM}_{tij2} \tag{4.2.24}$$

$$\xi^{M}_{tijkl} = \lambda^{M}_{\xi ijkl}\xi^{M}_{tijk} \qquad \forall\ k > 2 \tag{4.2.25}$$

$$\zeta^{UM}_{rtij2l} = \lambda^{UM}_{\zeta ij2l}\zeta^{UM}_{rtj2l} \tag{4.2.26}$$

$$\zeta^{CM}_{tij2l} = \lambda^{CM}_{\zeta ij2l}\zeta^{CM}_{tj2l} \tag{4.2.27}$$

$$\zeta^{M}_{tijkl} = \lambda^{M}_{\zeta ijkl}\zeta^{M}_{tjkl} \qquad \forall\ k > 2 \tag{4.2.28}$$

Remarks. The existence of the common factors ξ_{tij1}, ξ^{UM}_{rtij2}, ξ^{CM}_{tij2}, ξ^{M}_{tijk}, ζ^{UM}_{rtj2l}, ζ^{CM}_{tj2l}, and ζ^{M}_{tjkl} follows directly from Equations (4.2.14) - (4.2.20) of Assumptions (d) - (f) of Definition 4.1. Equation (4.2.22) follows from Equation (4.2.12) and (4.2.21) . Proofs of the existence of these latent variables were given by Koch (2013) and shall not be repeated here. Note that the constants $\alpha_{\xi ijkl}$ and $\lambda_{\xi ijkl}$, $k>1$, in Equation (4.2.22) differ from those in Equation (4.2.12) (α_{ijkl} and λ_{ijkl}), as they are a function of the parameters in Equation (4.2.12) and (4.2.21) ($\alpha_{ijkl}, \lambda_{ijkl}, \alpha_{\xi ij1l}$, and $\lambda_{\xi ij1l}$). Again, the parameters $\alpha_{\xi ijkl}$ and $\lambda_{\xi ijkl}$ and all of the coefficients κ_{sijkl} for the same i, j, k, and l are not separately identifiable (see Section 4.11 on identifiability conditions).
The term *common* refers to the fact that (1) the indicator-specific latent trait factors ξ_{tij1} and the latent trait method factors ξ^{UM}_{rtij2}, ξ^{CM}_{tij2} and ξ^{M}_{tijk} are common to all measurement occasions (occasion-unspecific factors), and (2) the occasion-specific state-residual method factors ζ^{UM}_{rtj2l}, ζ^{CM}_{tj2l} and ζ^{M}_{tjkl} are common to all indicators that belong to the same construct, the same method, and the same occasion of measurement (indicator-unspecific factors).

4.3 Uniqueness, admissible transformations and meaningful statements

From Theorem 4.1 it is apparent that the common latent trait variables ξ_{tij1}, common latent trait method variables ξ^{CM}_{tij2}, ξ^{UM}_{rtij2}, ξ^{M}_{tijk}, as well as the common latent state residual method variables ζ^{CM}_{tj2l}, ζ^{UM}_{tj2l}, and ζ^{M}_{tjkl} are not uniquely defined.
A detailed theorem and proofs on admissible transformations and uniqueness of the common latent (trait and state residual) method variables in the continuous-indicator LST-Com model with (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables are given by Koch (2013). These apply in the same manner to the LST-Com GRM. However, the uniqueness of the remaining parameters in the LST-Com GRM slightly differs from the continuous indicator model, so that a comprehensive theorem on the uniqueness of the latent variables in the LST-Com GRM is given in the following (cf., Eid, 1995; Koch, 2013).

Theorem 4.2. ***(Admissible transformations and uniqueness)***

1. Admissible Transformations
Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi, \boldsymbol{\lambda}^{\mathbf{M}}_\xi, \boldsymbol{\xi}^{\mathbf{UM}}_{rt}, \boldsymbol{\xi}^{\mathbf{CM}}_t, \boldsymbol{\xi}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LST-Com GRM with:*

$$\boldsymbol{\pi}_{rt} = (\pi_{rt1121}, \ldots, \pi_{rtij2l}, \ldots, \pi_{rtc_d d2f})^T \tag{4.3.1}$$

$$\boldsymbol{\pi}_t = (\pi_{t1111}, \ldots, \pi_{tijkl}, \ldots, \pi_{tc_d def})^T \qquad k \neq 2 \tag{4.3.2}$$

$$\boldsymbol{\kappa} = (\kappa_{11111}, \ldots, \kappa_{sijkl}, \ldots, \kappa_{(q_{c_d d}-1)c_d def})^T \tag{4.3.3}$$

$$\boldsymbol{\alpha}_\xi = (\alpha_{\xi 1111}, \ldots, \alpha_{\xi ijkl}, \ldots, \alpha_{\xi c_d def})^T \tag{4.3.4}$$
$$\boldsymbol{\lambda}_\xi = (\lambda_{\xi 1111}, \ldots, \lambda_{\xi ijkl}, \ldots, \lambda_{\xi c_d def})^T \tag{4.3.5}$$
$$\boldsymbol{\xi}_t = (\xi_{t111}, \ldots, \xi_{tij1}, \ldots, \xi_{tc_d d1})^T \tag{4.3.6}$$
$$\boldsymbol{\lambda}_\zeta = (\lambda_{\zeta 1121}, \ldots, \lambda_{\zeta ijkl}, \ldots, \lambda_{\zeta c_d def})^T \quad k > 1 \tag{4.3.7}$$
$$\boldsymbol{\zeta}_t = (\zeta_{t1111}, \ldots, \zeta_{tij1l}, \ldots, \zeta_{tc_d d1f})^T \tag{4.3.8}$$

$$\boldsymbol{\lambda}_\xi^{\mathbf{UM}} = (\lambda_{\xi 1121}^{\mathbf{UM}}, \ldots, \lambda_{\xi ij2l}^{\mathbf{UM}}, \ldots, \lambda_{\xi c_d d2f}^{\mathbf{UM}})^T \tag{4.3.9}$$
$$\boldsymbol{\lambda}_\xi^{\mathbf{CM}} = (\lambda_{\xi 1121}^{\mathbf{CM}}, \ldots, \lambda_{\xi ij2l}^{\mathbf{CM}}, \ldots, \lambda_{\xi c_d d2f}^{\mathbf{CM}})^T \tag{4.3.10}$$
$$\boldsymbol{\lambda}_\xi^{\mathbf{M}} = (\lambda_{\xi 1131}^{\mathbf{M}}, \ldots, \lambda_{\xi ijkl}^{\mathbf{M}}, \ldots, \lambda_{\xi c_d def}^{\mathbf{M}})^T \quad k > 2 \tag{4.3.11}$$
$$\boldsymbol{\xi}_{rt}^{\mathbf{UM}} = (\xi_{rt112}^{\mathbf{UM}}, \ldots, \xi_{rtij2}^{\mathbf{UM}}, \ldots, \xi_{rtc_d d2}^{\mathbf{UM}})^T \tag{4.3.12}$$
$$\boldsymbol{\xi}_t^{\mathbf{CM}} = (\xi_{t112}^{\mathbf{CM}}, \ldots, \xi_{tij2}^{\mathbf{CM}}, \ldots, \xi_{tc_d d2}^{\mathbf{CM}})^T \tag{4.3.13}$$
$$\boldsymbol{\xi}_t^{\mathbf{M}} = (\xi_{t113}^{\mathbf{M}}, \ldots, \xi_{tijk}^{\mathbf{M}}, \ldots, \xi_{tc_d de}^{\mathbf{M}})^T \quad k > 2 \tag{4.3.14}$$

$$\boldsymbol{\lambda}_\zeta^{\mathbf{UM}} = (\lambda_{\zeta 1121}^{\mathbf{UM}}, \ldots, \lambda_{\zeta ij2l}^{\mathbf{UM}}, \ldots, \lambda_{\zeta c_d d2f}^{\mathbf{UM}})^T \tag{4.3.15}$$
$$\boldsymbol{\lambda}_\zeta^{\mathbf{CM}} = (\lambda_{\zeta 1121}^{\mathbf{CM}}, \ldots, \lambda_{\zeta ij2l}^{\mathbf{CM}}, \ldots, \lambda_{\zeta c_d d2f}^{\mathbf{CM}})^T \tag{4.3.16}$$
$$\boldsymbol{\lambda}_\zeta^{\mathbf{M}} = (\lambda_{\zeta 1131}^{\mathbf{M}}, \ldots, \lambda_{\zeta ijkl}^{\mathbf{M}}, \ldots, \lambda_{\zeta c_d def}^{\mathbf{M}})^T \quad k > 2 \tag{4.3.17}$$
$$\boldsymbol{\zeta}_{rt}^{\mathbf{UM}} = (\zeta_{rt121}^{\mathbf{UM}}, \ldots, \zeta_{rtj2l}^{\mathbf{UM}}, \ldots, \zeta_{rtd2f}^{\mathbf{UM}})^T \tag{4.3.18}$$
$$\boldsymbol{\zeta}_t^{\mathbf{CM}} = (\zeta_{t121}^{\mathbf{CM}}, \ldots, \zeta_{tj2l}^{\mathbf{CM}}, \ldots, \zeta_{td2f}^{\mathbf{CM}})^T \tag{4.3.19}$$
$$\boldsymbol{\zeta}_t^{\mathbf{M}} = (\zeta_{t131}^{\mathbf{M}}, \ldots, \zeta_{tjkl}^{\mathbf{M}}, \ldots, \zeta_{tdef}^{\mathbf{M}})^T \quad k > 2 \tag{4.3.20}$$

If for all $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$:

$$\xi'_{tij1} = \beta_{\xi ij1}\xi_{tij1} + \gamma_{ij1} \tag{4.3.21}$$
$$\lambda'_{\xi ijkl} = \lambda_{\xi ijkl}/\beta_{\xi ij1} \tag{4.3.22}$$
$$\alpha'_{\xi ijkl} = \alpha_{\xi ijkl} - (\lambda_{\xi ijkl}/\beta_{\xi ij1})\gamma_{ij1} \tag{4.3.23}$$

$$\pi'_{tijkl} = \pi_{tijkl} + \nu_{ijkl} \quad k \neq 2 \tag{4.3.24}$$
$$\pi'_{rtij2l} = \pi_{rtij2l} + \nu_{ij2l} \tag{4.3.25}$$
$$\kappa'_{sijkl} = \kappa_{sijkl} + \nu_{ijkl} \tag{4.3.26}$$

$$\xi'^{\mathbf{UM}}_{rtij2} = \beta^{\mathbf{UM}}_{\xi ij2}\xi^{\mathbf{UM}}_{rtij2} \tag{4.3.27}$$
$$\xi'^{\mathbf{CM}}_{tij2} = \beta^{\mathbf{CM}}_{\xi ij2}\xi^{\mathbf{CM}}_{tij2} \tag{4.3.28}$$
$$\xi'^{\mathbf{M}}_{tijk} = \beta^{\mathbf{M}}_{\xi ijk}\xi^{\mathbf{M}}_{tijk} \quad k > 2 \tag{4.3.29}$$
$$\lambda'^{\mathbf{UM}}_{\xi ij2l} = \lambda^{\mathbf{UM}}_{\xi ij2l}/\beta^{\mathbf{UM}}_{\xi ij2} \tag{4.3.30}$$
$$\lambda'^{\mathbf{CM}}_{\xi ij2l} = \lambda^{\mathbf{CM}}_{\xi ij2l}/\beta^{\mathbf{CM}}_{\xi ij2} \tag{4.3.31}$$
$$\lambda'^{\mathbf{M}}_{\xi ijkl} = \lambda^{\mathbf{M}}_{\xi ijkl}/\beta^{\mathbf{M}}_{\xi ijk} \quad k > 2 \tag{4.3.32}$$

$$\zeta'^{\mathbf{UM}}_{rtj2l} = \beta^{\mathbf{UM}}_{\zeta j2l}\zeta^{\mathbf{UM}}_{rtj2l} \tag{4.3.33}$$
$$\zeta'^{\mathbf{CM}}_{tj2l} = \beta^{\mathbf{CM}}_{\zeta j2l}\zeta^{\mathbf{CM}}_{tj2l} \tag{4.3.34}$$
$$\zeta'^{\mathbf{M}}_{tjkl} = \beta^{\mathbf{M}}_{\zeta jkl}\zeta^{\mathbf{M}}_{tjkl} \qquad k > 2 \tag{4.3.35}$$
$$\lambda'^{\mathbf{UM}}_{\zeta ij2l} = \lambda^{\mathbf{UM}}_{\zeta ij2l}/\beta^{\mathbf{UM}}_{\zeta j2l} \tag{4.3.36}$$
$$\lambda'^{\mathbf{CM}}_{\zeta ij2l} = \lambda^{\mathbf{CM}}_{\zeta ij2l}/\beta^{\mathbf{CM}}_{\zeta j2l} \tag{4.3.37}$$
$$\lambda'^{\mathbf{M}}_{\zeta ijkl} = \lambda^{\mathbf{M}}_{\zeta ijkl}/\beta^{\mathbf{M}}_{\zeta jkl} \qquad k > 2 \tag{4.3.38}$$

where $\beta_{\xi ij1}$, $\beta^{\mathbf{UM}}_{\xi ij2}$, $\beta^{\mathbf{CM}}_{\xi ij2}$, $\beta^{\mathbf{M}}_{\xi ijk}$, $\beta^{\mathbf{UM}}_{\zeta j2l}$, $\beta^{\mathbf{CM}}_{\zeta j2l}$, $\beta^{\mathbf{M}}_{\zeta jkl} \in \mathbb{R}_+$, *and* γ_{ij1}, $\nu_{ijkl} \in \mathbb{R}$.
Then $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{rt}$, $\boldsymbol{\pi}'_{t}$, $\boldsymbol{\kappa}'$, $\boldsymbol{\alpha}'_{\xi}$, $\boldsymbol{\lambda}'_{\xi}$, $\boldsymbol{\xi}'_{t}$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_{t}$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_{\xi}$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_{\xi}$, $\boldsymbol{\lambda}'^{\mathbf{M}}_{\xi}$, $\boldsymbol{\xi}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}'^{\mathbf{CM}}_{t}$, $\boldsymbol{\xi}'^{\mathbf{M}}_{t}$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}'^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}'^{\mathbf{CM}}_{t}$, $\boldsymbol{\zeta}'^{\mathbf{M}}_{t}\rangle$ *is an LST-Com GRM, too, with*

$$\boldsymbol{\pi}'_{rt} = (\pi'_{rt1121}, \ldots, \pi'_{rtij2l}, \ldots, \pi'_{rtc_d d2f})^T \tag{4.3.39}$$
$$\boldsymbol{\pi}'_{t} = (\pi'_{t1111}, \ldots, \pi'_{tijkl}, \ldots, \pi'_{tc_d def})^T \qquad k \neq 2 \tag{4.3.40}$$
$$\boldsymbol{\kappa}' = (\kappa'_{11111}, \ldots, \kappa'_{sijkl}, \ldots, \kappa'_{(q_{c_d d}-1)c_d def})^T \tag{4.3.41}$$

$$\boldsymbol{\alpha}'_{\xi} = (\alpha'_{\xi 1111}, \ldots, \alpha'_{\xi ijkl}, \ldots, \alpha'_{\xi c_d def})^T \tag{4.3.42}$$
$$\boldsymbol{\lambda}'_{\xi} = (\lambda'_{\xi 1111}, \ldots, \lambda'_{\xi ijkl}, \ldots, \lambda'_{\xi c_d def})^T \tag{4.3.43}$$
$$\boldsymbol{\xi}'_{t} = (\xi'_{t111}, \ldots, \xi'_{tij1}, \ldots, \xi'_{tc_d d1})^T \tag{4.3.44}$$
$$\boldsymbol{\lambda}_{\zeta} = (\lambda_{\zeta 1121}, \ldots, \lambda_{\zeta ijkl}, \ldots, \lambda_{\zeta c_d def})^T \qquad k > 1 \tag{4.3.45}$$
$$\boldsymbol{\zeta}_{t} = (\zeta_{t1111}, \ldots, \zeta_{tij1l}, \ldots, \zeta_{tc_d d1f})^T \tag{4.3.46}$$

$$\boldsymbol{\lambda}'^{\mathbf{UM}}_{\xi} = (\lambda'^{\mathbf{UM}}_{\xi 1121}, \ldots, \lambda'^{\mathbf{UM}}_{\xi ij2l}, \ldots, \lambda'^{\mathbf{UM}}_{\xi c_d d2f})^T \tag{4.3.47}$$
$$\boldsymbol{\lambda}'^{\mathbf{CM}}_{\xi} = (\lambda'^{\mathbf{CM}}_{\xi 1121}, \ldots, \lambda'^{\mathbf{CM}}_{\xi ij2l}, \ldots, \lambda'^{\mathbf{CM}}_{\xi c_d d2f})^T \tag{4.3.48}$$
$$\boldsymbol{\lambda}'^{\mathbf{M}}_{\xi} = (\lambda'^{\mathbf{M}}_{\xi 1131}, \ldots, \lambda'^{\mathbf{M}}_{\xi ijkl}, \ldots, \lambda'^{\mathbf{M}}_{\xi c_d def})^T \qquad k > 2 \tag{4.3.49}$$
$$\boldsymbol{\xi}'^{\mathbf{UM}}_{rt} = (\xi'^{\mathbf{UM}}_{rt112}, \ldots, \xi'^{\mathbf{UM}}_{rtij2}, \ldots, \xi'^{\mathbf{UM}}_{rtc_d d2})^T \tag{4.3.50}$$
$$\boldsymbol{\xi}'^{\mathbf{CM}}_{t} = (\xi'^{\mathbf{CM}}_{t112}, \ldots, \xi'^{\mathbf{CM}}_{tij2}, \ldots, \xi'^{\mathbf{CM}}_{tc_d d2})^T \tag{4.3.51}$$
$$\boldsymbol{\xi}'^{\mathbf{M}}_{t} = (\xi'^{\mathbf{M}}_{t113}, \ldots, \xi'^{\mathbf{M}}_{tijk}, \ldots, \xi'^{\mathbf{M}}_{tc_d de})^T \qquad k > 2 \tag{4.3.52}$$

$$\boldsymbol{\lambda}'^{\mathbf{UM}}_{\zeta} = (\lambda'^{\mathbf{UM}}_{\zeta 1121}, \ldots, \lambda'^{\mathbf{UM}}_{\zeta ij2l}, \ldots, \lambda'^{\mathbf{UM}}_{\zeta c_d d2f})^T \tag{4.3.53}$$
$$\boldsymbol{\lambda}'^{\mathbf{CM}}_{\zeta} = (\lambda'^{\mathbf{CM}}_{\zeta 1121}, \ldots, \lambda'^{\mathbf{CM}}_{\zeta ij2l}, \ldots, \lambda'^{\mathbf{CM}}_{\zeta c_d d2f})^T \tag{4.3.54}$$
$$\boldsymbol{\lambda}'^{\mathbf{M}}_{\zeta} = (\lambda'^{\mathbf{M}}_{\zeta 1131}, \ldots, \lambda'^{\mathbf{M}}_{\zeta ijkl}, \ldots, \lambda'^{\mathbf{M}}_{\zeta c_d def})^T \qquad k > 2 \tag{4.3.55}$$
$$\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt} = (\zeta'^{\mathbf{UM}}_{rt121}, \ldots, \zeta'^{\mathbf{UM}}_{rtj2l}, \ldots, \zeta'^{\mathbf{UM}}_{rtd2f})^T \tag{4.3.56}$$
$$\boldsymbol{\zeta}'^{\mathbf{CM}}_{t} = (\zeta'^{\mathbf{CM}}_{t121}, \ldots, \zeta'^{\mathbf{CM}}_{tj2l}, \ldots, \zeta'^{\mathbf{CM}}_{td2f})^T \tag{4.3.57}$$
$$\boldsymbol{\zeta}'^{\mathbf{M}}_{t} = (\zeta'^{\mathbf{M}}_{t131}, \ldots, \zeta'^{\mathbf{M}}_{tjkl}, \ldots, \zeta'^{\mathbf{M}}_{tdef})^T \qquad k > 2 \tag{4.3.58}$$

2. Uniqueness

If both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_\xi$, $\boldsymbol{\lambda}_\xi$, $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}_\xi^{\mathbf{UM}}$, $\boldsymbol{\lambda}_\xi^{\mathbf{CM}}$, $\boldsymbol{\lambda}_\xi^{\mathbf{M}}$, $\boldsymbol{\xi}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\xi}_t^{\mathbf{CM}}$, $\boldsymbol{\xi}_t^{\mathbf{M}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{UM}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{CM}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{M}}$, $\boldsymbol{\zeta}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\zeta}_t^{\mathbf{CM}}$, $\boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ and $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{rt}$, $\boldsymbol{\pi}'_t$, $\boldsymbol{\kappa}'$, $\boldsymbol{\alpha}'_\xi$, $\boldsymbol{\lambda}'_\xi$, $\boldsymbol{\xi}'_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}'^{\mathbf{CM}}_t$, $\boldsymbol{\xi}'^{\mathbf{M}}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}'^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}'^{\mathbf{M}}_t\rangle$ are LST-Com GRMs, then for each $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$ there are γ_{ij1}, $\nu_{ijkl} \in \mathbb{R}$, and $\beta_{\xi ij1}$, $\beta_{\xi ij2}^{\mathbf{UM}}$, $\beta_{\xi ij2}^{\mathbf{CM}}$, $\beta_{\xi ijk}^{\mathbf{M}}$, $\beta_{\zeta j2l}^{\mathbf{UM}}$, $\beta_{\zeta j2l}^{\mathbf{CM}}$, $\beta_{\zeta jkl}^{\mathbf{M}} \in \mathbb{R}^+$, such that Equations (4.3.21) *to* (4.3.38) *hold.*

Remarks. Theorem 4.2 reveals that the the latent state residual variables ζ_{tij1l} as well as the loading parameters $\lambda_{\zeta ijkl}$ are uniquely defined in the LST-Com GRM with indicator-specific state residual variables ζ_{tij1l} (they are the same in $\mathcal{M}$ and $\mathcal{M}'$). This is the case as any translation of the latent response variables π_{tijkl} directly leads to the same translation for the latent trait variables $\xi_{tijkl} = \mathbb{E}[S_{tijkl} \mid p_T]$, with $S_{tijkl} = \pi_{tijkl}$, $k \neq 2$, and $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_t}]$, and thereby does not affect their residuals $\zeta_{tijkl} = S_{tijkl} - \xi_{tijkl}$. Therefore, there are no admissible transformations (except for the identity transformation) for these variables in this variant of the model, and meaningful statements can directly be made about the absolute values of ζ_{tij1l} and $\lambda_{\zeta ijkl}$. Note that this is not any longer true if the LST-Com GRM is defined with common latent state residual variables, as discussed in section 4.4.

Theorem 4.2 also shows that the common latent trait variables ξ_{tij1} are only uniquely defined up to linear transformations, while their loading parameters $\lambda_{\xi ijkl}$ are uniquely defined up to similarity transformations. The common trait method and common state residual method factors and their corresponding loading parameters are uniquely defined only up to similarity transformations. The parameters π'_{rtij2l}, π'_{tijkl}, $\alpha_{\xi ijkl}$, and κ_{sijkl} are uniquely defined up to translations by a constant. These properties and resulting meaningful statements are elaborated in the following. As these variables are not uniquely defined in the LST-Com GRM, a single representative of the set of possible values for these variables has to be chosen. This can be done by imposing different restrictions, which are described in section 4.11 on identifiability.

Common latent trait variables. To see that the common latent trait variables ξ_{tij1} are only uniquely defined up to linear transformations, let ξ'_{tij1}, $\lambda'_{\xi ij1l}$, and $\alpha'_{\xi ij1l}$ be defined as given by Equations (4.3.21) - (4.3.23). Then, it holds that:

$$\begin{aligned}\xi_{tij1l} &= \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\xi_{tij1} = \alpha_{\xi ij1l} - \frac{\lambda_{\xi ij1l}}{\beta_{\xi ij1}}\gamma_{ij1} + \frac{\lambda_{\xi ij1l}}{\beta_{\xi ij1}}\gamma_{ij1} + \frac{\lambda_{\xi ij1l}}{\beta_{\xi ij1}}\beta_{\xi ij1}\xi_{tij1} \\ &= \alpha_{\xi ij1l} - \frac{\lambda_{\xi ij1l}}{\beta_{\xi ij1}}\gamma_{ij1} + \frac{\lambda_{\xi ij1l}}{\beta_{\xi ij1}}(\beta_{\xi ij1}\xi_{tij1} + \gamma_{ij1}) = \alpha'_{\xi ij1l} + \lambda'_{\xi ij1l}\xi'_{tij1}\end{aligned}$$

Similarly it can be shown that, $\forall\, k > 1$, $\mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}] = \alpha_{\xi ijkl} + \lambda_{\xi ijkl}\xi_{tij1} = \alpha'_{\xi ijkl} + \lambda'_{\xi ijkl}\xi'_{tij1}$. For a proof on the uniqueness see Koch (2013).

As linear transformations are permissible transformations of the common latent trait variables ξ_{tij1}, meaningful statements regarding the values of these variables are statements on the ratio of differences. That is, if both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_\xi$, $\boldsymbol{\lambda}_\xi$, $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}_\xi^{\mathbf{UM}}$, $\boldsymbol{\lambda}_\xi^{\mathbf{CM}}$, $\boldsymbol{\lambda}_\xi^{\mathbf{M}}$, $\boldsymbol{\xi}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\xi}_t^{\mathbf{CM}}$, $\boldsymbol{\xi}_t^{\mathbf{M}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{UM}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{CM}}$, $\boldsymbol{\lambda}_\zeta^{\mathbf{M}}$, $\boldsymbol{\zeta}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\zeta}_t^{\mathbf{CM}}$, $\boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ and $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{rt}$, $\boldsymbol{\pi}'_t$, $\boldsymbol{\kappa}'$, $\boldsymbol{\alpha}'_\xi$, $\boldsymbol{\lambda}'_\xi$, $\boldsymbol{\xi}'_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}'^{\mathbf{CM}}_t$, $\boldsymbol{\xi}'^{\mathbf{M}}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}'^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}'^{\mathbf{M}}_t\rangle$ are LST-Com GRMs, then, for ω_1,ω_2, $\omega_3,\omega_4 \in \Omega$, $r \in R$, $t \in T$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, it holds that

$$\frac{\xi_{tij1}(\omega_1) - \xi_{tij1}(\omega_2)}{\xi_{tij1}(\omega_3) - \xi_{tij1}(\omega_4)} = \frac{\xi'_{tij1}(\omega_1) - \xi'_{tij1}(\omega_2)}{\xi'_{tij1}(\omega_3) - \xi'_{tij1}(\omega_4)}$$

Hence, statements regarding the ratios of differences of different persons' values on the common latent trait variables are meaningful (Koch, 2013). Statements on the absolute values of the common latent trait variables ξ_{tij1} or on ratios of the values themselves, in contrast, are not meaningful. That is, the common latent trait variables are measured on an interval scale. As linear transformations of variables do not have an influence on their correlations, all statements regarding the correlations between the common latent trait variables (of different indicators or constructs) are meaningful. So are statements on variance components (see Koch, 2013), as $\lambda^2_{\xi ijkl} Var(\xi_{tij1}) = \lambda'^2_{\xi ijkl} Var(\xi'_{tij1})$.
The loading parameters $\lambda_{\xi ijkl}$, in contrast, are measured on a ratio scale. That is, as they are uniquely defined up to similarity transformations, meaningful statements are statements regarding the ratio of $\lambda_{\xi ijkl}$ and $\lambda_{\xi ijk'l'}$ for $k,k' \in K$, $k = k'$ or $k \neq k'$ and $l,l' \in L$, $l = l'$ or $l \neq l'$, as

$$\frac{\lambda_{\xi ijkl}}{\lambda_{\xi ijk'l'}} = \frac{\lambda'_{\xi ijkl}}{\lambda'_{\xi ijk'l'}}$$

Latent response variables π_{tijkl} and π_{rtij2l}. The following proof shows that the latent response variables π_{tijkl}, $k \neq 2$, and π_{rtij2l} are uniquely defined only up to translations.

Proof. *Admissible transformations and uniqueness of the latent response variables π'_{tijkl} and π'_{rtij2l}.*

Let π'_{tijkl}, π'_{rtij2l}, and κ'_{sijkl} be defined as given by Equations (4.3.24) - (4.3.26). *Then, it holds that:*

$$\begin{aligned}\pi_{tsijkl} &= \pi_{tijkl} - \kappa_{sijkl} = (\pi_{tijkl} + \nu_{ijkl}) - (\kappa_{sijkl} + \nu_{ijkl}) \\ &= \pi'_{tijkl} - \kappa'_{sijkl} \qquad\qquad \forall\, k \neq 2\end{aligned}$$

and

$$\begin{aligned}\pi_{rtsij2l} &= \pi_{rtij2l} - \kappa_{sij2l} = (\pi_{rtij2l} + \nu_{ij2l}) - (\kappa_{sij2l} + \nu_{ij2l}) \\ &= \pi'_{rtij2l} - \kappa'_{sij2l}\end{aligned}$$

Let both $\mathcal{M} = \langle(\Omega,\mathcal{A},\mathcal{P}),\, \boldsymbol{\pi}_{rt},\, \boldsymbol{\pi}_t,\, \boldsymbol{\kappa},\, \boldsymbol{\alpha}_\xi,\, \boldsymbol{\lambda}_\xi,\, \boldsymbol{\xi}_t,\, \boldsymbol{\lambda}_\zeta,\, \boldsymbol{\zeta}_t,\, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi,\, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi,\, \boldsymbol{\lambda}^{\mathbf{M}}_\xi,\, \boldsymbol{\xi}^{\mathbf{UM}}_{rt},\, \boldsymbol{\xi}^{\mathbf{CM}}_t,\, \boldsymbol{\xi}^{\mathbf{M}}_t,\, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta,\, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta,\, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta,\, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt},\, \boldsymbol{\zeta}^{\mathbf{CM}}_t,\, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ and $\mathcal{M}' = \langle(\Omega,\mathcal{A},\mathcal{P}),\, \boldsymbol{\pi}'_{rt},\, \boldsymbol{\pi}'_t,\, \boldsymbol{\kappa}',\, \boldsymbol{\alpha}'_\xi,\, \boldsymbol{\lambda}'_\xi,\, \boldsymbol{\xi}'_t,\, \boldsymbol{\lambda}_\zeta,\, \boldsymbol{\zeta}_t,\, \boldsymbol{\lambda}'^{\mathbf{UM}}_\xi,\, \boldsymbol{\lambda}'^{\mathbf{CM}}_\xi,\, \boldsymbol{\lambda}'^{\mathbf{M}}_\xi,\, \boldsymbol{\xi}'^{\mathbf{UM}}_{rt},\, \boldsymbol{\xi}'^{\mathbf{CM}}_t,\, \boldsymbol{\xi}'^{\mathbf{M}}_t,\, \boldsymbol{\lambda}'^{\mathbf{UM}}_\zeta,\, \boldsymbol{\lambda}'^{\mathbf{CM}}_\zeta,\, \boldsymbol{\lambda}'^{\mathbf{M}}_\zeta,\, \boldsymbol{\zeta}'^{\mathbf{UM}}_{rt},\, \boldsymbol{\zeta}'^{\mathbf{CM}}_t,\, \boldsymbol{\zeta}'^{\mathbf{M}}_t\rangle$ be LST-Com GRMs. Then it has to hold that $\pi_{tijkl} - \kappa_{sijkl} = \pi'_{tijkl} - \kappa'_{sijkl}$ and $\pi_{rtij2l} - \kappa_{sij2l} = \pi'_{rtij2l} - \kappa'_{sij2l}$ for all $s \in S_{ij}$, $i \in I_j$, $j \in J$, $k \in K, k \neq 2$, and $l \in L$. It follows that $\pi'_{tijkl} = \pi_{tijkl} - \kappa_{sijkl} + \kappa'_{sijkl}$, $k \neq 2$, and $\pi'_{rtij2l} = \pi_{rtij2l} - \kappa_{sij2l} + \kappa'_{sij2l}$. As the difference $\kappa'_{sijkl} - \kappa_{sijkl}$ has to be the same over all $s \in S_{ij}$ for each $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, one can define ν_{ijkl} as $\nu_{ijkl} = \kappa'_{sijkl} - \kappa_{sijkl}$.

□

As the definition of the common latent response variables π_{tsij1l}, $\pi_{rtsij2l}$ and π_{tsijkl}, $k > 2$, in the LST-Com GRM is the same as in the LS-Com GRM, all statements concerning admissible transformations, uniqueness and meaningful statements regarding these variables are the same, too. To see which meaningful statements can be made regarding the latent response variables π_{tijkl} and π_{rtij2l} and the threshold parameters κ_{sijkl} see Equations (2.5.38) - (2.5.44) and the respective explanations in section 2.5. Note that with π_{tijkl} and π_{rtij2l} being uniquely defined only up to translations by a real constant ν_{ijkl}, also the latent variables $S_{tijkl} = \pi_{tijkl}$, $k \neq 2$, $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$ and $\xi_{tijkl} = \mathbb{E}[S_{tijkl} \mid p_T]$ are uniquely defined only up to the addition of the same constant ν_{ijkl}.

Latent (trait and residual state) method variables. Theorem 4.2 also reveals that the common trait method and common state residual method factors and their corresponding loading parameters are uniquely defined only up to similarity transformations. That is, admissible transformations of these factors and loadings are the multiplications with positive real numbers. As the multiplication with positive real numbers is an admissible transformation of the latent trait method factors $\xi^{\mathbf{UM}}_{rtij2}$, $\xi^{\mathbf{CM}}_{tij2}$, and $\xi^{\mathbf{M}}_{tijk}$, $k > 2$, and their corresponding factor loadings, statements regarding the absolute value of the parameters are not meaningful. The same holds for the latent state residual method factors $\zeta^{\mathbf{UM}}_{rtj2l}$, $\zeta^{\mathbf{CM}}_{tj2l}$, and $\zeta^{\mathbf{M}}_{tjkl}$, $k > 2$, and their corresponding factor loadings. Meaningful statements for these parameters are statements regarding the ratio of specific values of the factor loadings or the ratio of the values of the latent (trait or state residual) method factors (see Geiser, 2008; Koch, 2013). That is, the latent trait method and latent state residual method variables in the LST-Com GRM are measured on a ratio scale. Proofs of the uniqueness and meaningfulness for the latent (trait and residual state) method factors in the LST-Com model (applying to the LST-Com GRM, too) can be found in Koch (2013). Possible meaningful statements on the latent method factors as well as on their factor loadings shall only be shortly illustrated here with the example of the common latent unique method trait and unique method state residual factors and loadings.

Let both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_\xi$, $\boldsymbol{\lambda}_\xi$, $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}^{\mathbf{CM}}_t$, $\boldsymbol{\xi}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ and $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{rt}$, $\boldsymbol{\pi}'_t$, $\boldsymbol{\kappa}'$, $\boldsymbol{\alpha}'_\xi$, $\boldsymbol{\lambda}'_\xi$, $\boldsymbol{\xi}'_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}'^{\mathbf{CM}}_t$, $\boldsymbol{\xi}'^{\mathbf{M}}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}'^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}'^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}'^{\mathbf{M}}_t\rangle$ be LST-Com GRMs defined by Equations (4.3.1) to (4.3.58). Then, for $\omega_1, \omega_2 \in \Omega$, $r \in R$, $t \in T$, $i, i' \in I_j$, $j \in J$, $k \in K$, and $l, l' \in L$, it holds that

$$\frac{\lambda^{\mathbf{UM}}_{\xi ij2l}}{\lambda^{\mathbf{UM}}_{\xi ij2l'}} = \frac{\lambda'^{\mathbf{UM}}_{\xi ij2l}}{\lambda'^{\mathbf{UM}}_{\xi ij2l'}} \tag{4.3.59}$$

$$\frac{\xi^{\mathbf{UM}}_{rtij2}(\omega_1)}{\xi^{\mathbf{UM}}_{rtij2}(\omega_2)} = \frac{\xi'^{\mathbf{UM}}_{rtij2}(\omega_1)}{\xi'^{\mathbf{UM}}_{rtij2}(\omega_2)} \tag{4.3.60}$$

and

$$\frac{\lambda^{\mathbf{UM}}_{\zeta ij2l}}{\lambda^{\mathbf{UM}}_{\zeta i'j2l}} = \frac{\lambda'^{\mathbf{UM}}_{\zeta ij2l}}{\lambda'^{\mathbf{UM}}_{\zeta i'j2l}} \tag{4.3.61}$$

$$\frac{\zeta^{\mathbf{UM}}_{rtj2l}(\omega_1)}{\zeta^{\mathbf{UM}}_{rtj2l}(\omega_2)} = \frac{\zeta'^{\mathbf{UM}}_{rtj2l}(\omega_1)}{\zeta'^{\mathbf{UM}}_{rtj2l}(\omega_2)} \tag{4.3.62}$$

Thus, meaningful statements regarding the factor loadings of the common unique method trait variables are statements on the ratio of the loadings $\lambda^{\mathbf{UM}}_{\xi ij2l}$ belonging to the same indicator i of construct j but different measurement occasions l and l'. Meaningful statements regarding the factor loadings of the common unique method state residual variables, in contrast, are statements on the ratio of the loadings $\lambda^{\mathbf{UM}}_{\zeta ij2l}$ belonging to the same construct j and measurement occasion l, but different indicators i and i'. For the common trait or residual state unique method variables, statements on the ratio of two targets' values on the factors are meaningful. The values on the latent (trait or state residual) method factors of different targets can therefore be compared using their ratio. Meaningful statements with regard to ξ^{CM}_{tij2} and ξ^{M}_{tijk}, ζ^{CM}_{tj2l}, ζ^{M}_{tjkl}, and their respective loading parameters can be made in the same manner (Koch, 2013).

Note that comparisons of targets' values on the common latent trait method variables of different indicators $i \neq i'$ or constructs $j \neq j'$ are, as a consequence, also only meaningful if they refer to the ratio of the values of ω_1 and ω_2. The same holds for comparisons of the latent state residual method variables values of different constructs $j \neq j'$ or measurement occasions $l \neq l'$. Comparisons between the loading parameters of the latent state residual method factors of different measurement occasions

$l \neq l'$ are only meaningful if they refer to the ratios of loading parameters of different indicators $i \neq i'$ of the same construct and measurement occasion (see Equation 4.3.61). Analogously, comparisons between the loading parameters of the latent trait method variables of different indicators or constructs are only meaningful if they refer to the ratios of loading parameters of trait method variables of different measurement occasions $l \neq l'$ but the same indicator and construct (see Equation 4.3.59). Furthermore, the products $(\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2})$, $(\lambda^{CM}_{\zeta ij2l})^2 Var(\zeta^{CM}_{tj2l})$, $(\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2})$, $(\lambda^{UM}_{\zeta ij2l})^2 Var(\zeta^{UM}_{rtj2l})$, $(\lambda^{M}_{\xi ijkl})^2 Var(\xi^{M}_{tijk})$, and $(\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl})$ are invariant under similarity transformations, as, e.g.,

$$\begin{aligned}(\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2}) &= \frac{(\lambda^{UM}_{\xi ij2l})^2}{(\beta^{\mathbf{UM}}_{\xi ij2})^2}(\beta^{\mathbf{UM}}_{\xi ij2})^2 Var(\xi^{UM}_{rtij2}) \\ &= (\lambda'^{UM}_{\xi ij2l})^2 Var(\xi'^{UM}_{rtij2})\end{aligned} \tag{4.3.63}$$

(see Equation 4.3.30). Hence, any statement with respect to the ratio of variances are meaningful. Furthermore, statements concerning latent correlations between (trait or residual state) method factors are meaningful, as, for $i, i' \in I_j$ and $j, j' \in J$

$$Corr(\xi^{UM}_{rtij2}, \xi^{UM}_{rti'j'2}) = Corr(\xi'^{UM}_{rtij2}, \xi'^{UM}_{rti'j'2}) \tag{4.3.64}$$

and

$$Corr(\zeta^{UM}_{rtj2l}, \zeta^{UM}_{rtj'2l}) = Corr(\zeta'^{UM}_{rtj2l}, \zeta'^{UM}_{rtj'2l}) \tag{4.3.65}$$

hold by the general properties of correlations (Steyer & Nagel, 2017, remark 7.21, p. 243)). For a more detailed treatment and proofs of the meaningfulness of the latent (trait and residual state) method variables in the LST-Com models, see Koch (2013).

4.4 Common latent trait and state residual factors

The LST-Com GRM defined in section 4.2 and depicted in Figure 4.1 could also be defined with common latent state factors ζ_{tj1l} for all indicators belonging to the same construct j and measurement occasion l, instead of the indicator-specific latent state residual variables ζ_{tij1l}. The specification of common latent state residual factors is based on the assumption that the indicator-specific latent state factors ζ_{tij1l} and $\zeta_{ti'j1l}$ of two different indicators $i \neq i'$ pertaining to the same construct j and same occasion of measurement l, are similarity transformations of each other and therefore perfectly correlated, i.e., for each construct j, measured by the reference method ($k = 1$), and for each pair $(i, i') \in I_j \times I_j$, $(i \neq i')$, there are constants $\lambda_{\zeta ii'j1l} \in \mathbb{R}^+$ such that

$$\zeta_{tij1l} = \lambda_{\zeta ii'j1l}\zeta_{ti'j1l} \tag{4.4.1}$$

That is, a common latent state residual variable ζ_{tj1l} and constants $\lambda_{\zeta ij1l} \in \mathbb{R}^+$ exist so that the indicator-specific variables ζ_{itj1l} can be expressed by:

$$\zeta_{tij1l} = \lambda_{\zeta ij1l}\zeta_{tj1l} \tag{4.4.2}$$

The proof of the existence of these common latent state residual variables ζ_{tj1l} and their loading parameters $\lambda_{\zeta ij1l}$ follows the same logic as the proof for the existence of the common latent state residual method variables $\zeta^{\mathbf{UM}}_{rtj2l}$, $\zeta^{\mathbf{CM}}_{tj2l}$, and $\zeta^{\mathbf{M}}_{tjkl}$ (see Koch, 2013), and a proof of the existence of these variables in a multistate-multitrait model that is analogous to the present case can be found in Eid (1995). Note that the expectations of the latent state residual variables ζ_{tij1l} and ζ_{tj1l} are zero by definition and hence Equations (4.4.1) and (4.4.2) do not include intercepts. Note that with the definition of common latent state residual variables ζ_{tj1l} the regression of ζ_{tijkl} on the reference-method state residuals (given in Equation 4.2.13) changes, as they are now regressed on the common latent state residual variables ζ_{tjkl}. The resulting loading parameters $\lambda'_{\zeta ijkl}$ of the regression $\mathbb{E}[\zeta_{tijkl} \mid \zeta_{tj1l}] = \lambda'_{\zeta ijkl}\zeta_{tj1l}$ are given by

$$\lambda'_{\zeta ijkl} = \lambda_{\zeta ijkl}\lambda_{\zeta ij1l} \qquad k > 1 \tag{4.4.3}$$

The common latent state residual variables ζ_{tj1l} and their loading parameters $\lambda_{\zeta ij1l}$ and $\lambda'_{\zeta ijkl}$, $k > 1$, are not uniquely defined in the LST-Com GRM. Admissible transformations for these variables are multiplications with a positive real number, that is, they are uniquely defined up to similarity transformations. Hence, meaningful statements on the common latent state residual variables ζ_{tj1l} are statements regarding the ratio of different targets' values, as for all $t \in T$, $j \in J$, $l \in L$, and $\omega_1, \omega_2 \in \Omega$ it holds that

$$\frac{\zeta_{tj1l}(\omega_1)}{\zeta_{tj1l}(\omega_2)} = \frac{\zeta'_{tj1l}(\omega_1)}{\zeta'_{tj1l}(\omega_2)}. \tag{4.4.4}$$

with

$$\zeta'_{tj1l} = \beta_{\zeta j1l}\zeta_{tj1l} \tag{4.4.5}$$

and $\beta_{\zeta j1l} \in \mathbb{R}^+$. Meaningful statements on the new loading parameters $\lambda'_{\zeta ijkl}$ are statements regarding their ratio, that is, for all $i, i' \in I_j$, $i \neq i'$, $k, k' \in K$, $k \neq k'$, $j \in J$, and $l \in L$ it holds that

$$\frac{\lambda'_{\zeta ijkl}}{\lambda'_{\zeta i'jk'l}} = \frac{\lambda''_{\zeta ijkl}}{\lambda''_{\zeta i'jk'l}} \tag{4.4.6}$$

with

$$\lambda''_{\zeta ijkl} = \lambda'_{\zeta ijkl}/\beta_{\zeta j1l} \tag{4.4.7}$$

as

$$\begin{aligned}\frac{\lambda'_{\zeta ijkl}}{\lambda'_{\zeta i'jk'l}} &= \frac{\lambda'_{\zeta ijkl}/\beta_{\zeta j1l}}{\lambda'_{\zeta i'jk'l}/\beta_{\zeta j1l}} \\ &= \frac{\lambda''_{\zeta ijkl}}{\lambda''_{\zeta i'jk'l}}\end{aligned} \tag{4.4.8}$$

Additionally, it could be assumed that the latent trait variables ξ_{tij1} and $\xi_{ti'j1}$ of different indicators $i, i' \in I_j$, $i \neq i'$, belonging to the same construct j are linear transformations of each other (and hence are perfectly correlated). That is, it can be assumed that for each construct j, measured by the reference method ($k = 1$), and for each pair $(i, i') \in I_j \times I_j$, $(i \neq i')$, there are constants $\lambda_{ii'j1} \in \mathbb{R}^+$ and $\delta_{ii'j1} \in \mathbb{R}$ such that

$$\xi_{tij1} = \delta_{ii'j1} + \lambda_{ii'j1}\xi_{ti'j1}. \tag{4.4.9}$$

Again, this implies the existence of common latent trait variables ξ_{tj1} and constants $\lambda_{ij1} \in \mathbb{R}^+$ and $\delta_{ij1} \in \mathbb{R}$ such that

$$\xi_{tij1} = \delta_{ij1} + \lambda_{ij1}\xi_{tj1} \tag{4.4.10}$$

From Equation (4.4.10) it is obvious that the common latent trait variables ξ_{tj1} are only uniquely defined up to linear transformations, while the coefficients λ_{ij1} are uniquely defined up to similarity transformations and the coefficients δ_{ij1} up to translations by a real constant. Note that the coefficients $\lambda_{\xi ijkl}$ and $\alpha_{\xi ijkl}$ in the Equation for $\mathbb{E}[\xi_{tijkl} \mid \xi_{tij1l}]$ (see Equation 4.2.22) change from the model with indicator-specific latent trait variables ξ_{tij1} to the model with common latent trait variables ξ_{tj1}, as they are now the coefficients of the regression of the non-reference method variables ξ_{tijkl} on ξ_{tj1} (instead of on ξ_{tij1}). As the variables ξ_{tij1} as well as the coefficients $\lambda_{\xi ijkl}$ and $\alpha_{\xi ijkl}$ are measured on the same scale as their analogues in Equation (4.4.10), all the properties derived for these variables in section 4.3 apply to ξ_{tj1}, λ_{ij1} and δ_{ij1} in an analogous manner, respectively. Meaningful statements regarding the common latent trait factors ξ_{tj1} are therefore statements on the ratio of differences between different values of ξ_{tj1}, that is, for $\omega_1, \omega_2, \omega_3, \omega_4 \in \Omega$, $t \in T$, and $j \in J$ it holds that

$$\frac{\xi_{tj1}(\omega_1) - \xi_{tj1}(\omega_2)}{\xi_{tj1}(\omega_3) - \xi_{tj1}(\omega_4)} = \frac{\xi'_{tj1}(\omega_1) - \xi'_{tj1}(\omega_2)}{\xi'_{tj1}(\omega_3) - \xi'_{tj1}(\omega_4)} \tag{4.4.11}$$

For the new loading parameters $\lambda_{\xi ijkl}$ in an LST-Com GRM with common latent trait factors ξ_{tj1} meaningful statements are statements regarding the ratio of $\lambda_{\xi ijkl}$ and $\lambda_{\xi i'jk'l'}$ for $i,i' \in I_j$, $i = i'$ or $\neq i'$, $k,k' \in K$, $k = k'$ or $k \neq k'$ and $l,l' \in L$, $l = l'$ or $l \neq l'$, as

$$\frac{\lambda_{\xi ijkl}}{\lambda_{\xi i'jk'l'}} = \frac{\lambda'_{\xi ijkl}}{\lambda'_{\xi i'jk'l'}} \tag{4.4.12}$$

This result is a direct consequence of the definition of the variables analogous to the results in section 4.3 and the proof is left to the reader. Note that, as π_{tij1l} and κ_{sijkl} are only uniquely defined up to translations, the coefficients δ_{ij1}, $\alpha_{\xi ij1l}$ and all of the coefficients κ_{sij1l} for the same i, j, and l are not separately identifiable. The same holds for the coefficients $\alpha_{\xi ijkl}$ and κ_{sijkl}, $k > 1$. For further restrictions imposed on the mean structure, as well as the coefficients $\alpha_{\xi ijkl}$, δ_{ij1} and κ_{sijkl} due to identifiability considerations refer to Sections 4.10 and 4.11.
Also the indicator-specific common latent trait method factors $\xi^{\mathbf{UM}}_{rtij2}$, $\xi^{\mathbf{CM}}_{tij2}$, and $\xi^{\mathbf{M}}_{tijk}$, $k > 2$ for different indicators $i, i' \in I_j$, $i = i'$ but the same construct j could be assumed to be perfectly correlated, resulting in common latent trait method variables $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$, $k > 2$. Again, the variables $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$, $k > 2$, are uniquely defined only up to similarity transformations, and all properties discussed for the indicator-specific variables $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$ in section 4.2 hold for the variables $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$, $k > 2$, too. The same is true for their loading parameters $\lambda^{\mathbf{UM}}_{\xi ij2l}$, $\lambda^{\mathbf{CM}}_{\xi ij2l}$ and $\lambda^{\mathbf{M}}_{\xi ijkl}$, with the only difference that in an LST-Com GRM with common (non-indicator-specific) latent trait variables $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$, $k > 2$, meaningful statements on the loading parameters also refer to ratios of loading parameter for different indicators $i \neq i'$ of the same construct j. That is, for $i,i' \in I_j$, $j \in J$, $k \in K$, and $l, l' \in L$, it holds that

$$\frac{\lambda^{\mathbf{M}}_{\xi ijkl}}{\lambda^{\mathbf{M}}_{\xi i'jkl'}} = \frac{\lambda'^{\mathbf{M}}_{\xi ijkl}}{\lambda'^{\mathbf{M}}_{\xi i'jkl'}} \tag{4.4.13}$$

and analogously for the coefficients $\lambda^{\mathbf{UM}}_{\xi ij2l}$ and $\lambda^{\mathbf{CM}}_{\xi ij2l}$. An LST-Com GRM with common latent state residual factors ζ_{tj1l}, common latent trait factors ξ_{tj1}, and common latent trait method factors $\xi^{\mathbf{UM}}_{rtj2}$, $\xi^{\mathbf{CM}}_{tj2}$, and $\xi^{\mathbf{M}}_{tjk}$, $k > 2$, is depicted in Figure 4.2.

4.5 True score variables

The definition of latent true score variables for ordered categorical variables in the LST-Com GRM is identical to that of the LS-Com GRM and was given in Definition 2.2 in section 2.7.

4.6 Factor analytical representation

The LST-Com GRM presented above can also be represented as a factor model for ordinal data. As this representation does not depend on the specific model (LS-Com or LST-Com), the factor-analytical representation of the LST-Com GRM is identical to that of the LS-Com GRM as defined in section 2.8.

4.7 Independence assumptions and testability

4.7.1 LST-Com GRM with conditional independence

In order to derive testable consequences of the LST-Com GRM, several independence assumptions have to be introduced. These assumptions define the LST-Com GRM with conditional independence. Note that classical assumptions of multilevel modeling are made, that is, the targets are assumed to be independently and randomly drawn from a set of targets and the interchangeable raters are assumed to be independently and randomly drawn from a set of interchangeable raters given a target. The following assumptions extend these independence assumptions.

Definition 4.2. (LST-Com GRM with conditional independence)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_\xi$, $\boldsymbol{\lambda}_\xi$, $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}^{\mathbf{CM}}_t$, $\boldsymbol{\xi}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LST-Com GRM with* (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})*-congeneric variables.* $\mathcal{M}$ *is called LST-Com GRM with conditional independence if and only if the assumptions given in Definition 2.3 (i.e., Equations* (2.9.1) - (2.9.5)*), as well as the following assumption hold:*

$$p_R \perp\!\!\!\perp p_{TS_l} \mid p_T \tag{4.7.1}$$

Remarks. Assumptions (2.9.1) - (2.9.5) given in Definition 2.3 have the same meaning in the LST-Com GRM as in the LS-Com GRM and are explained in detail in the remarks to Definition 2.3. Assumption 4.7.1 states that, given the target, the sampling of the raters is independent of any target-situation realized on any of the measurement occasions. This assumption is not especially restrictive, as the interchangeable raters are supposed to be sampled randomly given the target and only once (at a time point different from any of the measurement occasions) and should thereby, given the target, not depend on specific situations realized for the target on any of the measurement occasions. Assumption 4.7.1 allows for the interpretation of the latent unique method trait variables ξ^{UM}_{rtij2l} as the difference between the conditional expectation of the latent response variables π_{rtij2l} given the target and the rater and its conditional expectation given the target only (see section 4.7.2). Furthermore, all of the conditional independence assumptions given in Definition 4.2 (also see Definition 2.3) imply consequences regarding the conditional and unconditional distributions of the observed variables Y_{tijkl} and Y_{rtij2l} as well as a specific covariance structure of the latent variables π_{tijkl} and π_{rtij2l} in the LST-Com GRM. Whether the restrictions imposed on the probability distributions of the response vectors and on the covariance structure by the conditional independence assumptions hold in empirical applications can be tested. That is, the conditional independence assumptions given in Definition 4.2 impose testable consequences on the covariance structure of the LST-Com GRM. These are derived in Section 4.8. The covariance structure of the variables Y^*_{tijkl} and Y^*_{rtij2l} is derived based on the covariance structure of the latent variables π_{tijkl} and π_{rtij2l} using the following theorem.

Theorem 4.3. ***(LST-Com GRM with conditional independence)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}_\xi^{\mathbf{UM}}, \boldsymbol{\lambda}_\xi^{\mathbf{CM}}, \boldsymbol{\lambda}_\xi^{\mathbf{M}}, \boldsymbol{\xi}_{rt}^{\mathbf{UM}}, \boldsymbol{\xi}_t^{\mathbf{CM}}, \boldsymbol{\xi}_t^{\mathbf{M}}, \boldsymbol{\lambda}_\zeta^{\mathbf{UM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{CM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{M}}, \boldsymbol{\zeta}_{rt}^{\mathbf{UM}}, \boldsymbol{\zeta}_t^{\mathbf{CM}}, \boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ *be an LST-Com GRM with* (ξ_{tij1l}, ξ_{tij2l}^{CM}, ξ_{rtij2l}^{UM}, ξ_{tijkl}^{M}, ζ_{tij2l}^{CM}, ζ_{tij2l}^{UM}, ζ_{tijkl}^{M})*-congeneric variables with conditional independence. Then, for all* $i, i' \in I_j$, $j, j' \in J$, $k \in K$, $l, l' \in L$, *and* y_{rtij2l}, $y_{tijkl} \in S_{ij}$ *it holds that:*

$$P\left(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}(Y_{rtij2l} = y_{rtij2l})\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\le e, k\neq 2}(Y_{tijkl} = y_{tijkl}) \,\middle|\, \pi_{t1111}, \ldots, \pi_{tc_d def}, \pi_{rt1121}, \ldots, \pi_{rtc_d d2f}\right) \tag{4.7.2}$$

$$= \prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}\prod_{k\le e, k\neq 2} P(Y_{tijkl} = y_{tijkl} \mid \pi_{tijkl}) \prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P(Y_{rtij2l} = y_{rtij2l} \mid \pi_{rtij2l})$$

Furthermore, it holds that:

$$P\left(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}(Y_{rtij2l} = y_{rtij2l})\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\le e, k\neq 2}(Y_{tijkl} = y_{tijkl}) \,\middle|\, \xi_{t111}, \ldots, \xi_{tc_d d1}, \zeta_{t1111}, \ldots, \zeta_{c_d d1f}, \xi_{t113}^{M}, \ldots, \xi_{tc_d de}^{M}, \zeta_{t131}^{M}, \ldots, \zeta_{tdef}^{M}, \xi_{t112}^{CM}, \ldots, \xi_{tc_d d2}^{CM}, \zeta_{t121}^{CM}, \ldots, \zeta_{td2f}^{CM}, \xi_{rt112}^{UM}, \ldots, \xi_{rtc_d d2}^{UM}, \zeta_{rt121}^{UM}, \ldots, \zeta_{rtd2f}^{UM}\right)$$

$$= \prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P(Y_{tij1l} = y_{tij1l} \mid \xi_{tij1}, \zeta_{tij1l}) \prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P(Y_{rtij2l} = y_{rtij2l} \mid \xi_{tij1}, \zeta_{tij1l}, \xi_{tij2}^{CM}, \zeta_{tj2l}^{CM}, \xi_{rtij2}^{UM}, \zeta_{rtj2l}^{UM}) \prod_{l=1}^{f}\prod_{k=3}^{e}\prod_{j=1}^{d}\prod_{i=1}^{c_j} P(Y_{tijkl} = y_{tijkl} \mid \xi_{tij1}, \zeta_{tij1l}, \xi_{tijk}^{M}, \zeta_{tjkl}^{M}) \tag{4.7.3}$$

Remarks. Equation (4.7.2) follows from Equations (2.9.1) and (2.9.2) - (2.9.4). This is the case as the random variables π_{tij1l}, π_{tijkl}, $k > 2$, and π_{rtij2l} are (p_T, p_{TS_l})-, (p_T, p_{TS_l}, $p_{R_kS_l}$)-, and (p_T, p_{TS_l}, p_R, $p_{R_2S_l}$)-measurable functions, respectively. Similar arguments lead to Equation (4.7.3). A proof was given by Eid (1995, pp. 97-98) for a comparable model and is applicable to the present case. According to Equation (4.7.2), all observed variables Y_{rtij2l} and Y_{tijkl} are independent given the latent response variables π_{rtij2l} and π_{tijkl}. Note that this assumption and its implications do not differ from the LS-Com GRM. Equation (4.7.2) implies that all associations between the observed variables are determined by the latent variables π_{tijkl} and π_{rtij2l} and their associations. According to Equation (4.7.3), the same holds with respect to the variables ξ_{tij1}, ζ_{tij1l}, ξ_{rtij2}^{UM}, ζ_{rtj2l}^{UM}, ξ_{tij2}^{CM}, ζ_{tj2l}^{CM}, ξ_{tijk}^{M}, and ζ_{tjkl}^{M}.

4.7.2 Conditional regressive independence of the latent state variables

As in the continuous-indicator LST-Com model, it can be shown that under conditional independence, the latent unique method trait variables ξ^{UM}_{rtij2l} can be interpreted as the difference between the conditional expectation of the latent response variables π_{rtij2l} given the target and the rater and its conditional expectation given the target only (Koch, 2013). The following theorem is a byproduct of the independence assumption given in Equation (4.7.1) of Definition 4.2.

Theorem 4.4. ***(LST-Com GRM with conditional regressive independent latent state variables)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi, \boldsymbol{\lambda}^{\mathbf{M}}_\xi, \boldsymbol{\xi}^{\mathbf{UM}}_{rt}, \boldsymbol{\xi}^{\mathbf{CM}}_t, \boldsymbol{\xi}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LST-Com GRM as defined by Definition 4.1.* $\mathcal{M}$ *is called LST-Com GRM with conditionally regressive independent* S_{tij2l} *variables, if assumption* (4.7.1) *of Definition 4.2 holds. Then, it follows that*

$$\mathbb{E}[S_{tij2l} \mid p_T, p_R] = \mathbb{E}[S_{tij2l} \mid p_T] \tag{4.7.4}$$

and the variables ξ^{UM}_{rtij2l} *can be redefined as follows:*

$$\xi^{UM}_{rtij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[\pi_{rtij2l} \mid p_T] \tag{4.7.5}$$

Remarks. Equation (4.7.4) states that the the expectation of the Level-2 variables S_{tij2l} (which are defined as $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$ in Equation 4.2.5) does not depend on the rater projection p_R given the target p_T. That is they are conditionally independent of p_R given p_T. Recall that the unique method variables UM_{rtij2l} represent the true deviation of a particular rater's rating from the expected rating over all interchangeable raters for target t on measurement occasion l (see Equation 4.2.1). The latent unique method trait variables ξ^{UM}_{rtij2l} are defined as the part of this deviation that does not depend on the situation variables p_{TS_l} and $p_{R_2S_l}$ (see Equation 4.2.2). Theorem 4.4 states that, given the assumption given in Equation (4.7.4), these variables can also be interpreted as the difference between the conditional expectation of the latent response variables π_{rtij2l} given the target and the rater and its conditional expectation given the target only. The result in Equation (4.7.5) of Theorem 4.4 is a direct consequence of assumption (4.7.4) and the definition of the latent variables given in Definition 4.1, as (cf. Koch, 2013):

$$\begin{aligned}
\xi^{UM}_{rtij2l} &= \mathbb{E}[UM_{rtij2l} \mid p_T, p_R] && \text{by Equation (4.2.2)}\\
&= \mathbb{E}[\pi_{rtij2l} - S_{tij2l} \mid p_T, p_R] && \text{by Equation (4.2.1) and (4.2.5)}\\
&= \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[S_{tij2l} \mid p_T, p_R] &&\\
&= \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[S_{tij2l} \mid p_T] && \text{by Equation (4.7.4)}\\
&= \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[\mathbb{E}(\pi_{rtij2l} \mid p_T, p_{TS_l}) \mid p_T] && \text{by Equation (4.2.5)}\\
&= \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[\pi_{rtij2l} \mid p_T] &&\\
&= \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \xi_{tij2l} && \text{by Equation (4.2.6)}
\end{aligned}$$

Hence, under this assumption, the latent unique method trait variables ξ^{UM}_{rtij2l} represent the over- or underestimation of a target's trait ξ_{tij2l} (as measured by all of the interchangeable raters per target) by a particular interchangeable rater r that does not depend on the specific (target- or rater-)situations realized on measurement occasion l.

4.7.3 LST-Com GRM in subpopulations

If the LST-Com GRM holds in a population, the model implies that it also holds in every subpopulation. That is, the item parameters α, λ, and κ have the same value in different subpopulations and the values on the latent variables π_{tijkl}, π_{rtij2l}, ξ_{tij1}, ζ_{tij1l}, ξ^{UM}_{rtij2}, ζ^{UM}_{rtj2l}, ξ^{CM}_{tij2}, ζ^{CM}_{tj2l}, ξ^{M}_{tijk}, and ζ^{M}_{tjkl} remain the same when considering subpopulations, given that the parameterization and scaling of the latent variables is the same. This fact was proven by Eid (1995, pp. 94-96, 99) for a comparable model. The prove applies to the present model, too, and shall therefore not be repeated here.

Furthermore, if an LST-Com GRM with conditional independence holds in a population, the same conditional independence assumptions also hold in subpopulations. That is, the covariance structure implied by the conditional independence assumptions has to hold in every subpopulation. For a prove for a comparable model see Eid (1995). While the covariance structure has to be the same in every subpopulation, the values of the (non-zero) variances and covariances between the latent variables are allowed to vary between subpopulations.

4.8 Covariance structure

The LST-Com GRM with conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}. The following theorem introduces the covariances that are zero as a result of the conditional independence assumptions.

Theorem 4.5. ***(Testability)***

If $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_\xi$, $\boldsymbol{\lambda}_\xi$, $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\xi$, $\boldsymbol{\lambda}^{\mathbf{M}}_\xi$, $\boldsymbol{\xi}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\xi}^{\mathbf{CM}}_t$, $\boldsymbol{\xi}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ is an LST-Com GRM with (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables and conditional independence, then, for all $i, i' \in I_j$, $j, j' \in J$, $k \in K$, and $l, l' \in L$, with $i = i'$ or $i \neq i'$, $j = j'$ or $j \neq j'$, and $l = l'$ or $l \neq l'$, it holds that

1. *The latent trait variables are uncorrelated with the latent trait method variables:*

$$Cov(\xi_{tij1}, \xi^{CM}_{tij2}) = 0 \tag{4.8.1}$$
$$Cov(\xi_{tij1}, \xi^{UM}_{rti'j'2}) = 0 \tag{4.8.2}$$
$$Cov(\xi_{tij1}, \xi^{M}_{tijk}) = 0 \qquad k > 2 \tag{4.8.3}$$

2. *The latent state residual variables are uncorrelated with the latent state residual method variables:*

$$Cov(\zeta_{tij1l}, \zeta^{CM}_{tj2l}) = 0 \tag{4.8.4}$$
$$Cov(\zeta_{tij1l}, \zeta^{UM}_{rtj'2l'}) = 0 \tag{4.8.5}$$
$$Cov(\zeta_{tij1l}, \zeta^{M}_{tjkl}) = 0 \qquad k > 2 \tag{4.8.6}$$

3. *The latent trait variables are uncorrelated with all latent state residual (method) variables:*

$$Cov(\xi_{tij1}, \zeta_{ti'j'kl'}) = 0 \tag{4.8.7}$$
$$Cov(\xi_{tij1}, \zeta^{CM}_{tj'2l'}) = 0 \tag{4.8.8}$$
$$Cov(\xi_{tij1}, \zeta^{UM}_{rtj'2l'}) = 0 \tag{4.8.9}$$
$$Cov(\xi_{tij1}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \tag{4.8.10}$$

4. *The latent trait method variables are uncorrelated with all latent state residual (method) variables:*

$$Cov(\xi^{CM}_{tij2}, \zeta_{ti'j'kl'}) = 0 \tag{4.8.11}$$
$$Cov(\xi^{CM}_{tij2}, \zeta^{CM}_{tj'2l'}) = 0 \tag{4.8.12}$$
$$Cov(\xi^{CM}_{tij2}, \zeta^{UM}_{rtj'2l'}) = 0 \tag{4.8.13}$$
$$Cov(\xi^{CM}_{tij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \tag{4.8.14}$$

$$Cov(\xi^{UM}_{rtij2}, \zeta_{ti'j'kl'}) = 0 \tag{4.8.15}$$
$$Cov(\xi^{UM}_{rtij2}, \zeta^{CM}_{tj'2l'}) = 0 \tag{4.8.16}$$
$$Cov(\xi^{UM}_{rtij2}, \zeta^{UM}_{rtj'2l'}) = 0 \tag{4.8.17}$$
$$Cov(\xi^{UM}_{rtij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \tag{4.8.18}$$

$$Cov(\xi^{M}_{tijk}, \zeta_{ti'j'kl'}) = 0 \qquad k > 2 \tag{4.8.19}$$
$$Cov(\xi^{M}_{tijk}, \zeta^{CM}_{tj'2l'}) = 0 \qquad k > 2 \tag{4.8.20}$$
$$Cov(\xi^{M}_{tijk}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad k > 2 \tag{4.8.21}$$
$$Cov(\xi^{M}_{tijk}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \tag{4.8.22}$$

5. *Uncorrelatedness of latent trait method variables:*

$$Cov(\xi^{CM}_{tij2}, \xi^{UM}_{rti'j'2}) = 0 \tag{4.8.23}$$
$$Cov(\xi^{M}_{tijk}, \xi^{UM}_{rti'j'2}) = 0 \qquad k > 2 \tag{4.8.24}$$

6. *Uncorrelatedness of latent state residual method variables:*

$$Cov(\zeta^{CM}_{tj2l}, \zeta^{UM}_{rtj'2l'}) = 0 \tag{4.8.25}$$
$$Cov(\zeta^{M}_{tjkl}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad k > 2 \tag{4.8.26}$$

Proofs.
The following proofs are based on Definitions 4.1 and 4.2 (and thereby Definition 2.3), as well as on general properties of residual variables. These properties are that residual variables are always uncorrelated with their regressors as well as with measurable functions of their regressors (Steyer & Nagel, 2017, p. 323). Some of these proofs have already been pre-published for the continuous-indicator model in Koch et al. (2017)[1].

4.8.1 By Equation (4.2.24), *the latent common method trait factor* ξ^{CM}_{tij2} *can be rewritten as* $\xi^{CM}_{tij2} =$

[1] This concerns proofs [4.8.2], [4.8.6], [4.8.7], [4.8.9], [4.8.22], and [4.8.24].

$\xi^{CM}_{tij2l}/\lambda^{CM}_{\xi ij2l}$. ξ^{CM}_{tij2l} *is defined as* $\xi^{CM}_{tij2l} = \xi_{tij2l} - \mathbb{E}[\xi_{tij2l} \mid \xi_{tij1l}]$ *by Equation* (4.2.8). *Hence,* ξ^{CM}_{tij2l} *is defined as a residual with respect to* ξ_{tij1l}. *As the latent trait variable* ξ_{tij1l} *can be rewritten as* $\xi_{tij1l} = \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\xi_{tij1}$ *(see Equation 4.2.21),* ξ^{CM}_{tij2l} *is also defined as a residual with respect to* ξ_{tij1}. *As residuals are uncorrelated with their regressors, it follows that, for the same indicator i and construct j,* $Cov(\xi_{tij1}, \xi^{CM}_{tij2}) = 0$.

4.8.2 By Equation (4.2.23), *the latent unique method trait factor* $\xi^{UM}_{rti'j'2}$ *can be rewritten as* $\xi^{UM}_{rti'j'2l'}/\lambda^{UM}_{\xi i'j'2l'}$. $\xi^{UM}_{rti'j'2l'}$ *is defined as* $\xi^{UM}_{rti'j'2l'} = \mathbb{E}[\pi_{rti'j'2l'} \mid p_T, p_R] - \mathbb{E}[\pi_{rti'j'2l'} \mid p_T]$ *by Equation* (4.7.5). *Hence,* $\xi^{UM}_{rti'j'2l'}$ *is defined as a residual with respect to a* p_T*-measurable function and thereby uncorrelated with any* p_T*-measurable function. It follows that* $\xi^{UM}_{rti'j'2l'}$ *is uncorrelated with the* p_T*-measurable function* $\xi_{tij1} = (\xi_{tij1l} - \alpha_{\xi ij1l})/\lambda_{\xi ij1l}$.

4.8.3 The proof of Equation (4.8.3) *follows the same logic as Proof 4.8.1.*

4.8.4 By equation (4.2.27) ζ^{CM}_{tj2l} *can be rewritten as* $\zeta^{CM}_{tj2l} = \zeta^{CM}_{tij2l}/\lambda^{CM}_{\zeta ij2l}$. *Hence* $Cov(\zeta_{tij1l}, \zeta^{CM}_{tj2l}) = 0 \iff Cov(\zeta_{tij1l}, \zeta^{CM}_{tij2l}) = 0$. ζ^{CM}_{tij2l} *is defined as* $\zeta_{tij2l} - \mathbb{E}[\zeta_{tij2l} \mid \zeta_{tij1l}]$ *by Equation* (4.2.10). *That is,* ζ^{CM}_{tij2l} *is defined as a residual with respect to* ζ_{tij1l}. *As residuals are uncorrelated with their regressors, it follows that, for the same indicator i, construct j, and measurement occasion l,* $Cov(\zeta_{tij1l}, \zeta^{CM}_{tij2l}) = 0$.

4.8.5 By Equation (4.2.26) $\zeta^{UM}_{rtj'2l'}$ *can be rewritten as* $\zeta^{UM}_{rti'j'2l'}/\lambda^{UM}_{\zeta i'j'2l'}$. *Hence* $Cov(\zeta_{tij1l}, \zeta^{UM}_{rtj'2l'}) = 0 \iff Cov(\zeta_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0$. $\zeta^{UM}_{rti'j'2l'}$ *is defined as* $UM_{rt'i'j2l'} - \xi^{UM}_{rti'j'2l'}$ *by Equation* (4.2.3). *Hence* $Cov(\zeta_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0 \iff Cov(\zeta_{tij1l}, UM_{rti'j'2l'}) = 0$ *and* $Cov(\zeta_{tij1l}, \xi^{UM}_{rti'j'2l'}) = 0$. $Cov(\zeta_{tij1l}, \xi^{UM}_{rti'j'2l'}) = 0$ *is shown in Proof 4.8.15.*
$Cov(\zeta_{tij1l}, UM_{rti'j'2l'}) = 0$ *holds if* $Cov(S_{tij1l}, UM_{rti'j'2l'}) = 0$ *and* $Cov(\xi_{tij1l}, UM_{rti'j'2l'}) = 0$. $Cov(S_{tij1l}, UM_{rti'j'2l'}) = 0$ *is shown in Proof 2.10.2.* $Cov(\xi_{tij1l}, UM_{rti'j'2l'}) = 0$ *holds as* ξ_{tij1l} *is a direct function of* S_{tij1l}, *and* $UM_{rti'j'2l'}$ *is a residual with respect to* S_{tij1l} *(see Proof 2.10.2).*

4.8.6 The proof of Equation (4.8.6) *follows the same logic as Proof 4.8.4.*

4.8.7 By Equation (4.2.21) ξ_{tij1} *can be rewritten as* $\xi_{tij1} = (\xi_{tij1l} - \alpha_{\xi ij1l})/\lambda_{\xi ij1l}$. *Hence* $Cov(\xi_{tij1}, \zeta_{ti'j'kl'}) = 0 \iff Cov(\xi_{tij1l}, \zeta_{ti'j'kl'}) = 0$. $\zeta_{ti'j'kl'}$ *is defined as* $\zeta_{ti'j'kl'} = S_{ti'j'kl'} - \mathbb{E}[S_{ti'j'kl'} \mid p_T]$ *by Equations* (4.2.7) *and* (4.2.6). *That is,* $\zeta_{ti'j'kl'}$ *is defined as residual with respect to any* p_T*-measurable function, and is therefore uncorrelated with the* p_T*-measurable function* ξ_{tij1l}.

4.8.8 Again, ξ_{tij1} *can be rewritten as* $(\xi_{tij1l} - \alpha_{\xi ij1l})/\lambda_{\xi ij1l}$. *By equation* (4.2.27) $\zeta^{CM}_{tj'2l'}$ *can be rewritten as* $\zeta^{CM}_{ti'j'2l'}/\lambda^{CM}_{\zeta i'j'2l'}$. *Hence* $Cov(\xi_{tij1}, \zeta^{CM}_{tj'2l'}) = 0 \iff Cov(\xi_{tij1l}, \zeta^{CM}_{ti'j'2l'}) = 0$. *By equation* (4.2.10) $\zeta^{CM}_{ti'j'2l'}$ *is defined as* $\zeta_{ti'j'2l'} - \mathbb{E}[\zeta_{ti'j'2l'} \mid \zeta_{ti'j'1l'}]$. *It follows that* $Cov(\xi_{tij1l}, \zeta^{CM}_{ti'j'2l'}) = 0$, *as* $Cov(\xi_{tij1l}, \zeta_{ti'j'2l'}) = 0$ *and* $Cov(\xi_{tij1l}, \zeta_{ti'j'1l'}) = 0$ *as shown in Proof 4.8.7.*

4.8.9 Again, ξ_{tij1} *can be rewritten as* $(\xi_{tij1l} - \alpha_{\xi ij1l})/\lambda_{\xi ij1l}$. *By equation* (4.2.26) $\zeta^{UM}_{rtj'2l'}$ *can be rewritten as* $\zeta^{UM}_{rti'j'2l'}/\lambda^{UM}_{\zeta i'j'2l'}$. *Hence* $Cov(\xi_{tij1}, \zeta^{UM}_{rtj'2l'}) = 0 \iff Cov(\xi_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0$. *By Equations* (4.2.3) *and* (4.2.2) $\zeta^{UM}_{rti'j'2l'}$ *is defined as* $UM_{rti'j'2l'} - \mathbb{E}[UM_{rti'j'2l'} \mid p_T, p_R]$. *That is,* $\zeta^{UM}_{rti'j'2l'}$ *is a residual with respect to* p_T*-measurable functions and thereby uncorrelated with the* p_T*-measurable function* ξ_{tij1l}.

4.8.10 The proof of Equation (4.8.10) *follows the same logic as Proof 4.8.8.*

4.8.11 ξ^{CM}_{tij2} *can be rewritten as* $\xi^{CM}_{tij2l}/\lambda^{CM}_{\xi ij2l}$ *by Equation* (4.2.24). *Hence,* $Cov(\xi^{CM}_{tij2}, \zeta_{ti'j'kl'}) = 0 \iff Cov(\xi^{CM}_{tij2l}, \zeta_{ti'j'kl'}) = 0$. ξ^{CM}_{tij2l} *is defined as* $\xi_{tij2l} - \mathbb{E}[\xi_{tij2l} \mid \xi_{tij1l}]$ *by Equation* (4.2.8).

That is, ξ^{CM}_{tij2l} is a direct function of the p_T-measurable functions ξ_{tij2l} and ξ_{tij1l}. As $\zeta_{ti'j'kl'}$ is defined as a residual with respect to and thereby uncorrelated with any p_T-measurable function (see Proof 4.8.7), it follows that $Cov(\xi^{CM}_{tij2l}, \zeta_{ti'j'kl'}) = 0$.

4.8.12 *$\zeta^{CM}_{tj'2l'}$ can be rewritten as $\zeta^{CM}_{ti'j'2l'}/\lambda^{CM}_{\zeta i'j'2l'}$ by Equation (4.2.27). $\zeta^{CM}_{ti'j'2l'}$ is defined as $\zeta_{ti'j'2l'} - \mathbb{E}[\zeta_{ti'j'2l'} \mid \zeta_{ti'j'1l'}]$ by Equation (4.2.10) and thereby a direct function of $\zeta_{ti'j'2l'}$ and $\zeta_{ti'j'1l'}$. As shown in Proof 4.8.11, $Cov(\xi^{CM}_{tj2l}, \zeta_{ti'j'kl'}) = 0\ \forall k$. It follows that $Cov(\xi^{CM}_{tij2}, \zeta^{CM}_{tj'2l'}) = 0$.*

4.8.13 *As argued in Proof 4.8.11, ξ^{CM}_{tij2} is a direct function of the p_T-measurable functions ξ_{tij2l} and ξ_{tij1l}. As argued in Proof 4.8.9, $\zeta^{UM}_{rti'j'2l'}$ is a residual with respect to p_T-measurable functions and thereby uncorrelated with the p_T-measurable functions ξ_{tij1l} and ξ_{tij2l}. It follows that $Cov(\xi^{CM}_{tij2}, \zeta^{UM}_{rtj'2l'}) = 0$.*

4.8.14 *The proof of Equation* (4.8.14) *follows the same logic as Proof 4.8.12.*

4.8.15 *By Equation* (4.2.23)*, the latent unique method trait factor ξ^{UM}_{rtij2} can be rewritten as $\xi^{UM}_{rtij2l}/\lambda^{UM}_{\xi ij2l}$. ξ^{UM}_{rtij2l} is defined as $\mathbb{E}[UM_{rtij2l} \mid p_T, p_R]$ by Equation* (4.2.2)*, that is, it is a (p_T, p_R)-measurable function. $\zeta_{ti'j'kl'}$ is defined as $\zeta_{ti'j'kl'} = S_{ti'j'kl'} - \mathbb{E}[S_{ti'j'kl'} \mid p_T]$ by Equations* (4.2.7) *and* (4.2.6)*. That is, $\zeta_{ti'j'kl'}$ is a (p_T, p_{TS_l})-measurable function, as $S_{ti'j'kl'}$ is a (p_T, p_{TS_l})-measurable function (see Equations 4.2.4 and 4.2.5). Furthermore, $\zeta_{ti'j'kl'}$ is defined as residual with respect to any p_T-measurable function. Hence, $\zeta_{ti'j'kl'}$ is uncorrelated with the (p_T, p_R)-measurable function ξ^{UM}_{rtij2l}, as $p_{TS_l} \perp\!\!\!\perp p_R \mid p_T$ by conditional independence assumption* (4.7.1).

4.8.16 *$\zeta^{CM}_{tj'2l'}$ can be rewritten as $\zeta^{CM}_{ti'j'2l'}/\lambda^{CM}_{\zeta i'j'2l'}$ by Equation (4.2.27). $\zeta^{CM}_{ti'j'2l'}$ is defined as $\zeta_{ti'j'2l'} - \mathbb{E}[\zeta_{ti'j'2l'} \mid \zeta_{ti'j'1l'}]$ by Equation (4.2.10) and thereby a direct function of $\zeta_{ti'j'2l'}$ and $\zeta_{ti'j'1l'}$. As shown in Proof 4.8.15, $Cov(\xi^{UM}_{rti'j'2l}, \zeta_{ti'j'kl'}) = 0\ \forall k$. It follows that $Cov(\xi^{UM}_{rtij2}, \zeta^{CM}_{tj'2l'}) = 0$.*

4.8.17 *By Equation* (4.2.23)*, the latent unique method trait factor $\xi^{UM}_{rti'j'2}$ can be rewritten as $\xi^{UM}_{rti'j'2} = \xi^{UM}_{rti'j'2l}/\lambda^{UM}_{\xi i'j'2l}$. By equation* (4.2.26) *$\zeta^{UM}_{rtj'2l'}$ can be rewritten as $\zeta^{UM}_{rti'j'2l'}/\lambda^{UM}_{\zeta i'j'2l'}$. Hence, $Cov(\xi^{UM}_{rtij2}, \zeta^{UM}_{rtj'2l'}) = 0 \iff Cov(\xi^{UM}_{rtij2l}, \zeta^{UM}_{rti'j'2l'}) = 0$. $\xi^{UM}_{rti'j'2l}$ is defined as $\mathbb{E}[UM_{rti'j'2l'} \mid p_T, p_R]$ by Equation* (4.2.2)*. By Equations* (4.2.3) *and* (4.2.2) *$\zeta^{UM}_{rti'j'2l'}$ is defined as $UM_{rti'j'2l'} - \mathbb{E}[UM_{rti'j'2l'} \mid p_T, p_R]$. That is, $\zeta^{UM}_{rti'j'2l'}$ is a residual with respect to (p_T, p_R)-measurable functions and thereby uncorrelated with the (p_T, p_R)-measurable function ξ^{UM}_{rtij2l}.*

4.8.18 *The proof of Equation* (4.8.18) *follows the same logic as Proof 4.8.16.*

4.8.19 *The proof of Equation* (4.8.19) *follows the same logic as Proof 4.8.11.*

4.8.20 *The proof of Equation* (4.8.20) *follows the same logic as Proof 4.8.12.*

4.8.21 *The proof of Equation* (4.8.21) *follows the same logic as Proof 4.8.13.*

4.8.22 *The proof of Equation* (4.8.22) *follows the same logic as Proof 4.8.12.*

4.8.23 *$Cov(\xi^{CM}_{tij2}, \xi^{UM}_{rti'j'2}) = 0 \iff Cov(\xi^{CM}_{tij2l}, \xi^{UM}_{rti'j'2l'}) = 0$, as ξ^{CM}_{tij2} can be rewritten as $\xi^{CM}_{tij2l}/\lambda^{CM}_{\xi ij2l}$ by Equation* (4.2.24) *and $\xi^{UM}_{rti'j'2}$ can be rewritten as $\xi^{UM}_{rti'j'2l'}/\lambda^{UM}_{\xi i'j'2l'}$ by Equation* (4.2.23)*. ξ^{CM}_{tij2l} is defined as $\xi_{tij2l} - \mathbb{E}[\xi_{tij2l} \mid \xi_{tij1l}]$ and thereby a direct function of the p_T-measurable functions ξ_{tij2l} and ξ_{tij1l}. As shown in Proof 4.8.2, $\xi^{UM}_{rti'j'2l'}$ is uncorrelated with ξ_{tij2l} and ξ_{tij1l}, as it is defined as a residual with respect to any p_T-measurable function. It follows that $Cov(\xi^{CM}_{tij2l}, \xi^{UM}_{rti'j'2l'}) = 0$.*

4.8.24 *The proof of Equation* (4.8.24) *follows the same logic as Proof 4.8.23.*

4.8.25 *ζ^{CM}_{tj2l} can be rewritten as $\zeta^{CM}_{tij2l}/\lambda^{CM}_{\zeta ij2l}$ by Equation (4.2.27). ζ^{CM}_{tij2l} is defined as $\zeta_{tij2l} - \mathbb{E}[\zeta_{tij2l} \mid \zeta_{tij1l}]$ by Equation (4.2.10) and thereby a direct function of ζ_{tij2l} and ζ_{tij1l}. $Cov(\zeta_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0$ is shown in Proof 4.8.5. As ζ_{tij2l} also is a (p_T, p_{TS_l})-measurable function (as is ζ_{tij1l}; see Equations 4.2.4 and 4.2.5), Proof 4.8.5 applies to ζ_{tij2l} in an analogous manner. It follows that $Cov(\zeta^{CM}_{tj2l}, \zeta^{UM}_{rtj'2l'}) = 0$.*

4.8.26 *By Equation (4.2.28), ζ^{M}_{tjkl} can be rewritten as $\zeta^{M}_{tijkl}/\lambda^{M}_{\zeta ijkl}$. ζ^{M}_{tijkl} is defined as $\zeta_{tijkl} - \mathbb{E}[\zeta_{tijkl} \mid \zeta_{tij1l}]$, $k > 2$, by Equation (4.2.11) and thereby a direct function of ζ_{tijkl}, $k > 2$, and ζ_{tij1l}. By Equation (4.2.26) $\zeta^{UM}_{rtj'2l'}$ can be rewritten as $\zeta^{UM}_{rti'j'2l'}/\lambda^{UM}_{\zeta i'j'2l'}$. Hence, for $k > 2$:*

$$Cov(\zeta^{M}_{tjkl}, \zeta^{UM}_{rtj'2l'}) = 0 \iff Cov(\zeta_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0 \text{ and } Cov(\zeta_{tijkl}, \zeta^{UM}_{rti'j'2l'}) = 0.$$

$Cov(\zeta_{tij1l}, \zeta^{UM}_{rti'j'2l'}) = 0$ is shown in Proof 4.8.5.

$\zeta^{UM}_{rti'j'2l'}$ is defined as $UM_{rt'i'j2l'} - \xi^{UM}_{rti'j'2l'}$ by Equation (4.2.3). Hence, for $k > 2$, $Cov(\zeta_{tijkl}, \zeta^{UM}_{rti'j'2l'}) = 0$ holds if $Cov(\zeta_{tijkl}, UM_{rti'j'2l'}) = 0$ and $Cov(\zeta_{tijkl}, \xi^{UM}_{rti'j'2l'}) = 0$ hold.

$Cov(\zeta_{tijkl}, \xi^{UM}_{rti'j'2l'}) = 0$, $\forall$ k, is shown in Proof 4.8.15.

$Cov(\zeta_{tijkl}, UM_{rti'j'2l'}) = 0$ holds if $Cov(S_{tijkl}, UM_{rti'j'2l'}) = 0$ and $Cov(\xi_{tijkl}, UM_{rti'j'2l'}) = 0$, as $\zeta_{tijkl} = S_{tijkl} - \xi_{tijkl}$. The latent state variable S_{tijkl} is defined as $S_{tijkl} = \pi_{tijkl}$, which is a $(p_T, p_{TS_l}, p_{R_kS_l})$-measurable function as it can be defined as $\pi_{tijkl} : \varphi_{tijkl}(p_T, p_{TS_l}, p_{R_kS_l})$, with $\varphi_{tijkl} : \Omega_T \times \Omega_{TS_l} \times \Omega_{R_kS_l} \to \mathbb{R}$ and $(p_T, p_{TS_l}, p_{R_kS_l}) : \Omega \to \Omega_T \times \Omega_{TS_l} \times \Omega_{R_kS_l}$ (see remarks to Definition 2.1 in Section 2.4).

$UM_{rti'j'2l'}$ is defined as $\pi_{rti'j'2l'} - \mathbb{E}[\pi_{rti'j'2l'} \mid p_T, p_{TS_{l'}}]$ by Equation (4.2.1). By conditional independence Assumption (2.9.5) the expression $\mathbb{E}(\pi_{rti'j'2l'} \mid p_T, p_{TS_{l'}})$ can be replaced by $\mathbb{E}(\pi_{rti'j'2l'} \mid p_T, p_{TS_1}, \ldots, p_{TS_{l'}}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_{l'}}, \ldots, p_{R_kS_f})$. Therefore, it follows that $UM_{rti'j'2l'}$ is also a residual with respect to a $(p_T, p_{TS_1}, \ldots, p_{TS_l}, \ldots, p_{TS_f}, p_{R_kS_1}, \ldots, p_{R_kS_l}, \ldots, p_{R_kS_f})$-measurable function, and thereby uncorrelated to the $(p_T, p_{TS_l}, p_{R_kS_l})$-measurable function S_{tijkl}, $k > 2$. It follows that $Cov(\xi_{tijkl}, UM_{rti'j'2l'}) = 0$ as ξ_{tijkl} is a direct function of S_{tijkl}, and $UM_{rti'j'2l'}$ is a residual with respect to S_{tijkl}.

Remarks. Note that the conditional independence assumptions given in Definition 4.2 do not imply that the latent state residuals ζ_{tij1l} or the latent state residual method factors ζ^{CM}_{tj2l}, ζ^{UM}_{rtj2l}, and ζ^{M}_{tjkl} are uncorrelated over time. That is, the LST-Com GRM as defined by Definitions 4.1 and 4.2 allows ζ_{tij1l} and $\zeta_{tij1l'}$, ζ^{CM}_{tj2l} and $\zeta^{CM}_{tj2l'}$, ζ^{UM}_{rtj2l} and $\zeta^{UM}_{rtj2l'}$, as well as ζ^{M}_{tjkl} and $\zeta^{CM}_{tj2l'}$ to be correlated for $l \neq l'$, respectively. Note that the fact that these correlations are not restricted to zero by definition of the LST-Com GRM, does not mean that they cannot be zero in empirical applications. In fact, it is necessary to set some or all of these correlations to zero for identification reasons. Nevertheless, the model in its current form does for example allow to model autoregressive processes on the level of the state residual (method) variables.

In the current definition of the LST-Com GRM, the following additional conditional independence assumption would lead to uncorrelated latent state residual and latent state residual method variables over time (Koch et al., 2017).

Definition 4.3. (LST-Com GRM with strong conditional independence)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}_\xi^{\mathbf{UM}}, \boldsymbol{\lambda}_\xi^{\mathbf{CM}}, \boldsymbol{\lambda}_\xi^{\mathbf{M}}, \boldsymbol{\xi}_{rt}^{\mathbf{UM}}, \boldsymbol{\xi}_t^{\mathbf{CM}}, \boldsymbol{\xi}_t^{\mathbf{M}}, \boldsymbol{\lambda}_\zeta^{\mathbf{UM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{CM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{M}}, \boldsymbol{\zeta}_{rt}^{\mathbf{UM}}, \boldsymbol{\zeta}_t^{\mathbf{CM}}, \boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ *be an LST-Com GRM with* (ξ_{tij1l}, ξ_{tij2l}^{CM}, ξ_{rtij2l}^{UM}, ξ_{tijkl}^{M}, ζ_{tij2l}^{CM}, ζ_{tij2l}^{UM}, ζ_{tijkl}^{M})*-congeneric variables.* $\mathcal{M}$ *is called LST-Com GRM with strong conditional independence if and only if the assumptions given in Definition 4.2 (i.e., LST-Com GRM with conditional independence) hold and the following statements hold:*

$$\begin{aligned}&\mathbb{E}[S_{tijkl} \mid p_T, p_{TS_1}, \dots, p_{TS_{l-1}}, p_{TS_{l+1}}, \dots, p_{TS_f}, p_{R_{k'}S_1}, \dots, p_{R_{k'}S_{l-1}}, p_{R_{k'}S_{l+1}}, \dots, p_{R_{k'}S_f}] \\ &\quad = \mathbb{E}[S_{tijkl} \mid p_T] \qquad \text{for } k = k' \text{ or } k \neq k',\ k' > 1\end{aligned} \tag{4.8.27}$$

$$\begin{aligned}&\mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_1}, \dots, p_{TS_{l-1}}, p_{TS_{l+1}}, \dots, p_{TS_f}, p_R, p_{R_2S_1}, \dots, p_{R_2S_{l-1}}, p_{R_2S_{l+1}}, \dots, p_{R_2S_f}] \\ &\quad = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R]\end{aligned} \tag{4.8.28}$$

Assumption (4.8.27) states that given the target variable p_T the Level-2 latent state variables S_{tijkl} measured on occasion l neither depend on target situations nor on rater-situations realized on different occasions of measurement l'. Similarly, Assumption (4.8.28) states that given the target p_T and interchangeable rater p_R the level-1 latent response variables π_{rtij2l} do not depend on target situations $p_{TS_{l'}}$ or rater situations $p_{R_2S_{l'}}$ of different measurement occasions $l' \neq l$. These two assumptions imply that occasion-specific residual (method) variables belonging to different occasions of measurement l and l' are uncorrelated with each other.

Theorem 4.6. ***(Testability of LST-Com GRM with strong conditional independence)***

If $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}_\xi^{\mathbf{UM}}, \boldsymbol{\lambda}_\xi^{\mathbf{CM}}, \boldsymbol{\lambda}_\xi^{\mathbf{M}}, \boldsymbol{\xi}_{rt}^{\mathbf{UM}}, \boldsymbol{\xi}_t^{\mathbf{CM}}, \boldsymbol{\xi}_t^{\mathbf{M}}, \boldsymbol{\lambda}_\zeta^{\mathbf{UM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{CM}}, \boldsymbol{\lambda}_\zeta^{\mathbf{M}}, \boldsymbol{\zeta}_{rt}^{\mathbf{UM}}, \boldsymbol{\zeta}_t^{\mathbf{CM}}, \boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ *is an LST-Com GRM with* (ξ_{tij1l}, ξ_{tij2l}^{CM}, ξ_{rtij2l}^{UM}, ξ_{tijkl}^{M}, ζ_{tij2l}^{CM}, ζ_{tij2l}^{UM}, ζ_{tijkl}^{M})*-congeneric variables and strong conditional independence, then Equations* (4.8.1) - (4.8.26) *of Theorem 4.5 hold and, for all* $i \in I_j$, $j \in J$, $k \in K$, *and* $l, l' \in L$, *it holds that the latent state residual (method) variables are uncorrelated over measurement occasions:*

$$Cov(\zeta_{tij1l}, \zeta_{tij1l'}) = 0 \tag{4.8.29}$$

$$Cov(\zeta_{tj2l}^{CM}, \zeta_{tj2l'}^{CM}) = 0 \tag{4.8.30}$$

$$Cov(\zeta_{tjkl}^{M}, \zeta_{tjkl'}^{M}) = 0 \qquad k > 2 \tag{4.8.31}$$

$$Cov(\zeta_{rtj2l}^{UM}, \zeta_{rtj2l'}^{UM}) = 0 \qquad k > 2 \tag{4.8.32}$$

$$Cov(\zeta_{tij1l}, \zeta_{tj2l'}^{CM}) = 0 \tag{4.8.33}$$

$$Cov(\zeta_{tij1l}, \zeta_{tjkl'}^{M}) = 0 \qquad k > 2 \tag{4.8.34}$$

$$Cov(\zeta_{tjkl}^{M}, \zeta_{tj2l'}^{CM}) = 0 \qquad k > 2 \tag{4.8.35}$$

Proofs. Note that some of these proofs have already been pre-published for the continuous-indicator model in Koch et al. (2017)[2].

4.8.29 By Equations (4.2.6) *and* (4.2.7), ζ_{tijkl} *is defined as* $S_{tijkl} - \mathbb{E}[S_{tijkl} \mid p_T]$.
According to Assumption (4.8.27), $\mathbb{E}[S_{tijkl} \mid p_T]$ *can be replaced by*
$\mathbb{E}[S_{tijkl} \mid p_T, p_{TS_1}, \ldots, p_{TS_{l-1}}, p_{TS_{l+1}}, \ldots, p_{TS_f}, p_{R_{k'}S_1}, \ldots, p_{R_{k'}S_{l-1}}, p_{R_{k'}S_{l+1}}, \ldots, p_{R_{k'}S_f}]$,
for $k = k'$ *or* $k \neq k'$, $k' > 1$.
That is, ζ_{tijkl} *is also a residual with respect to* $(p_T, p_{TS_{l'}})$*- and* $(p_T, p_{TS_{l'}}, p_{R_{k'}S_{l'}})$*-measurable functions,* $k = k'$ *or* $k \neq k'$, $k' > 1$. *It follows that* $Cov(\zeta_{tij1l}, \zeta_{tij1l'}) = 0$, $Cov(\zeta_{tijkl}, \zeta_{tijkl'}) = 0$ *and* $Cov(\zeta_{tij1l}, \zeta_{tijkl'}) = 0 \; \forall \; k$.

4.8.30 By equation (4.2.27) ζ^{CM}_{tj2l} *can be rewritten as* $\zeta^{CM}_{tj2l} = \zeta^{CM}_{tij2l} / \lambda^{CM}_{\zeta ij2l}$. ζ^{CM}_{tij2l} *is defined as* $\zeta_{tij2l} - \mathbb{E}[\zeta_{tij2l} \mid \zeta_{tij1l}]$ *by Equation* (4.2.10). *That is,* ζ^{CM}_{tj2l} *is a direct function of* ζ_{tij2l} *and* ζ_{tij1l}. *As shown in Proof 4.8.29, by Assumption* (4.8.27) *it holds that* $Cov(\zeta_{tijkl}, \zeta_{tijkl'}) = 0$ *and* $Cov(\zeta_{tij1l}, \zeta_{tijkl'}) = 0 \; \forall \; k$. *It follows that* $Cov(\zeta^{CM}_{tj2l}, \zeta^{CM}_{tj2l'}) = 0$.

4.8.31 The proof of Equation (4.8.31) *follows the same logic as Proof 4.8.30.*

4.8.32 By equation (4.2.26) ζ^{UM}_{rtj2l} *can be rewritten as* $\zeta^{UM}_{rtij2l} / \lambda^{UM}_{\zeta ij2l}$. *By Equations* (4.2.3) *and* (4.2.2) ζ^{UM}_{rtij2l} *is defined as* $UM_{rtij2l} - \mathbb{E}[UM_{rtij2l} \mid p_T, p_R]$, *with* $\mathbb{E}[UM_{rtij2l} \mid p_T, p_R] = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[\pi_{rtij2l} \mid p_T]$ *by Equation* (4.7.5).
According to assumption (4.8.28), $\mathbb{E}[\pi_{rtij2l} \mid p_T, p_R]$ *can be replaced by*
$\mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_1}, \ldots, p_{TS_{l-1}}, p_{TS_{l+1}}, \ldots, p_{TS_f}, p_R, p_{R_2S_1}, \ldots, p_{R_2S_{l-1}}, p_{R_2S_{l+1}}, \ldots, p_{R_2S_f}]$
That is, ζ^{UM}_{rtij2l} *is also a residual with respect to* $(p_T, p_{TS_{l'}}, p_R, p_{R_1S_{l'}})$*-measurable functions. It follows that* ζ^{UM}_{rtj2l} *is uncorrelated with the* $(p_T, p_{TS_{l'}}, p_R, p_{R_1S_{l'}})$*-measurable function* $\zeta^{UM}_{rtj2l'}$.

4.8.33 - 4.8.35 The proofs of Equation (4.8.33) - (4.8.35) *follow the same logic as Proofs 4.8.29 and 4.8.30.*

As shown above, Assumptions (4.8.27) and (4.8.28) imply that all of the stability over time is accounted for by the latent trait and latent trait method variables. Note that these assumptions might be violated in empirical applications and can thereby be a source of misfit (Bishop, Geiser, & Cole, 2015; Courvoisier, Eid, Lischetzke, & Schreiber, 2010; Eid, Courvoisier, & Lischetzke, 2012).
This might for example be the case if associations between measurements (in this case, the latent response variables) of adjacent occasions are the result of an autoregressive process [first order autoregressive process, AR(1)]. That is, not only the expectations of the probability distributions of the variables S_{tijkl} correlate over time (correlations between ξ_{tijkl} and $\xi_{tijkl'}$), but also the specific situations realized on adjacent measurement occasions are not independent. These kind of dependencies are most likely to occur in the case of short time-intervals between adjacent measurement occasions (e.g., several measurements during one day, as often found in ambulatory assessment data), due to "carry-over" effects (Bishop et al., 2015; Eid et al., 2012).
From a theoretical standpoint, depending on the measurement design and the construct under investigation, it seems appropriate to incorporate autoregressive effects to model short-term stability and account for the possibility that correlations between measurements decrease with increasing time-intervals. Many psychological constructs show decreasing stabilities with increasing time-intervals between measurements, while stabilities do not approach zero even over long time intervals (long-term stability; Cole et al., 2005; Prenoveau, 2016; Roberts & DelVecchio, 2000). Including autoregressive (residual) components in LST or latent growth-curve models has been found to adequately

[2]This concerns proofs [4.8.29] and [4.8.34].

reflect the covariance structure and increase model fit in longitudinal studies investigating different psychological phenomena (Bollen & Curran, 2004; Cole, 2006; Courvoisier et al., 2010; Eid et al., 2012; Lucas & Donnellan, 2007; Luhmann et al., 2011). On the other hand, underspecification due to non-modeled auto-regressive processes present in the data was found to bias latent growth curve model parameters (Ferron, Dailey, & Yi, 2002; Kwok, West, & Green, 2007; Sivo, Fan, & Witta, 2005). Furthermore, many methodological approaches to modeling longitudinal psychological data acknowledge the importance of combining autoregressive effects with long-lasting stabilities or change (Cole et al., 2005; Eid et al., 2012; Hamaker et al., 2015; Hamaker, 2005; Kenny & Zautra, 2001).

The LST-Com GRM with conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}, including the zero-correlations specified in Theorem 4.5. Whether this covariance structure holds in empirical applications can be tested based on the covariance structure of the variables Y^*_{tijkl} and Y^*_{rtij2l}, as defined in Section 2.8, with SEMs for ordinal observed variables. As derived in section 2.10, Equation (4.7.2) implies that $Cov(Y^*_{tijkl}, Y^*_{t(ijkl)'}) = Cov(\pi_{tijkl}, \pi_{t(ijkl)'})$ and $Cov(Y^*_{rtij2l}, Y^*_{rt(ij2l)'}) = Cov(\pi_{rtij2l}, \pi_{rt(ij2l)'})$, that is, all associations between the observed variables Y_{rtij2l} and Y_{tijkl} are determined by the latent variables π_{tijkl} and π_{rtij2l} and their associations. It follows that $Cov(\varepsilon^*_{tijkl}, \pi_{t(ijkl)'}) = 0$ and $Cov(\varepsilon^*_{tijkl}, \varepsilon^*_{t(ijkl)'}) = 0$ for all $i, i' \in I_j$, $j, j' \in J$, $k, k' \in K$, and $l, l' \in L$. The same applies to the Y^*_{rtij2l} and their residuals and the combination of Y^*_{rtij2l} and Y^*_{tijkl} and their residuals. As the residuals ε^*_{tijkl} and ε^*_{rtij2l} have to be uncorrelated with all π_{tijkl} and π_{rtij2l}, the residual variables are as well uncorrelated with all latent variables ξ_{tij1}, ξ^{CM}_{tij2}, ξ^{UM}_{rtij2}, ξ^{M}_{tijk}, ζ_{ij1l}, ζ^{CM}_{tj2l}, ζ^{UM}_{tj2l}, ζ^{M}_{tjkl}. This fact also follows from Equation (4.7.3). For a more detailed treatment see section 2.10.

The zero-correlations of the error variables with all other error variables and latent variables of the LST-Com GRM combined with the covariance structure of the latent response variables π_{tijkl} and π_{rtij2l} define the covariance structure of the variables Y^*_{rtij2l} and Y^*_{tijkl} in the LST-Com GRM. This covariance structure equals the covariance structure of the latent variables in the LST-Com model for continuous indicator variables derived by Koch (2013), with one exception. While the variance of the error variables is a variable that is free to vary and is estimated in the SEM with continuous indicator variables, this variance is fixed to one for ε^*_{tijkl} and ε^*_{rtij2l} for all $j \in J$, $i \in I_j$, $k \in K$, and $l \in L$ in the LST-Com GRM. This restriction guarantees the equivalence of the LST-Com GRM and the factor analytical representation of the model.

The total covariance matrix Σ_T of the variables Y^*_{rtij2l} and Y^*_{tijkl} in an LST-Com GRM with strong conditional independence can be partitioned, just as in the continuous case, into a within and a between covariance matrix and can be represented as

$$\Sigma_T = \Lambda_{\xi B}\Phi_{\xi B}\Lambda'_{\xi B} + \Lambda_{\zeta B}\Phi_{\zeta B}\Lambda'_{\zeta B} + \Theta_B + \Lambda_{\xi W}\Phi_{\xi W}\Lambda'_{\xi W} + \Lambda_{\zeta W}\Phi_{\zeta W}\Lambda'_{\zeta W} + \Theta_W \tag{4.8.36}$$

where $\Lambda_{\xi B}$ and $\Lambda_{\xi W}$ refer to the factor loading matrices of the trait-specific variables on the between- and within-level, respectively, $\Lambda_{\zeta B}$ and $\Lambda_{\zeta W}$ refer to the factor loading matrices of the occasion-specific variables on the between- and within-level, respectively, $\Phi_{\xi B}$ and $\Phi_{\xi W}$ refer to the variance-covariance matrices of the between and within trait-specific latent variables, respectively, $\Phi_{\zeta B}$ and $\Phi_{\zeta W}$ refer to the variance-covariance matrices of the between and within occasion-specific latent variables, respectively, and Θ_B and Θ_W are the between- and within-level residual variance-covariance matrices, where all non-zero elements $Var(E_{rtij2l})$ and $Var(E_{tijkl})$ in the matrices Θ_B and Θ_W have to be replaced by 1. For a detailed illustration of the covariance matrices and their elements see the supplementary material of Koch et al. (2017) or Koch (2013, pp. 121-127), for a detailed description of the interpretation of all the non-zero covariances and correlations in the LST-Com model refer to Koch (2013, pp. 127-129).

Table 4.1: Definition of the Consistency, Occasion-Specificity and different Method Specificity coefficients in the LST-Com GRM

Consistency and Method Specificity Coefficients		
Coefficient and method	Level	Definition
Time consistencies		
Self	Target	$Con(\pi_{tij1l}) = \frac{(\lambda_{\xi_{ij1l}})^2 Var(\xi_{tij1})}{Var(\pi_{tij1l})}$
Interchangeable	Rater	$Con(\pi_{rtij2l}) = \frac{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})+(\lambda^{UM}_{\xi_{ij2l}})^2 Var(\xi^{UM}_{rtij2})}{Var(\pi_{rtij2l})}$
Structurally different	Target	$Con(\pi_{tijkl}) = \frac{(\lambda_{\xi_{ijkl}})^2 Var(\xi_{tij1})+(\lambda^{M}_{\xi_{ijkl}})^2 Var(\xi^{M}_{tijk})}{Var(\pi_{tijkl})}$
Occasion Specificities		
Self	Target	$OSpe(\pi_{tij1l}) = \frac{Var(\zeta_{tij1l})}{Var(\pi_{tij1l})}$
Interchangeable	Rater	$OSpe(\pi_{rtij2l}) = \frac{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})+(\lambda^{UM}_{\zeta_{ij2l}})^2 Var(\zeta^{UM}_{rtj2l})}{Var(\pi_{rtij2l})}$
Structurally different	Target	$OSpe(\pi_{tijkl}) = \frac{(\lambda_{\zeta_{ijkl}})^2 Var(\zeta_{tij1l})+(\lambda^{M}_{\zeta_{ijkl}})^2 Var(\zeta^{M}_{tjkl})}{Var(\pi_{tijkl})}$
Trait method consistencies		
Interchangeable	Rater	$TCon(\pi_{rtij2l}) = \frac{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})}{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})+(\lambda^{UM}_{\xi_{ij2l}})^2 Var(\xi^{UM}_{rtij2})}$
Interchangeable	Target	$TCon(\pi_{tij2l}) = \frac{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})}{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})}$
Structurally different	Target	$TCon(\pi_{tijkl}) = \frac{(\lambda_{\xi_{ijkl}})^2 Var(\xi_{tij1})}{(\lambda_{\xi_{ijkl}})^2 Var(\xi_{tij1})+(\lambda^{M}_{\xi_{ijkl}})^2 Var(\xi^{M}_{tijk})}$
Trait method specificities		
Interchangeable	Rater	$TUMS(\pi_{rtij2l}) = \frac{(\lambda^{UM}_{\xi_{ij2l}})^2 Var(\xi^{UM}_{rtij2})}{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})+(\lambda^{UM}_{\xi_{ij2l}})^2 Var(\xi^{UM}_{rtij2})}$
Interchangeable	Rater	$TCMS(\pi_{rtij2l}) = \frac{(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})}{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})+(\lambda^{UM}_{\xi_{ij2l}})^2 Var(\xi^{UM}_{rtij2})}$
Interchangeable	Target	$TCMS(\pi_{tij2l}) = \frac{(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})}{(\lambda_{\xi_{ij2l}})^2 Var(\xi_{tij1})+(\lambda^{CM}_{\xi_{ij2l}})^2 Var(\xi^{CM}_{tij2})}$
Structurally different	Target	$TMS(\pi_{tijkl}) = \frac{(\lambda^{M}_{\xi_{ijkl}})^2 Var(\xi^{M}_{tijk})}{(\lambda_{\xi_{ijkl}})^2 Var(\xi_{tij1})+(\lambda^{M}_{\xi_{ijkl}})^2 Var(\xi^{M}_{tijk})}$
Occasion-specific method consistencies		
Interchangeable	Rater	$OCon(\pi_{rtij2l}) = \frac{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})}{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})+(\lambda^{UM}_{\zeta_{ij2l}})^2 Var(\zeta^{UM}_{rtj2l})}$
Interchangeable	Target	$OCon(\pi_{tij2l}) = \frac{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})}{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})}$
Structurally different	Target	$OCon(\pi_{tijkl}) = \frac{(\lambda_{\zeta_{ijkl}})^2 Var(\zeta_{tij1l})}{(\lambda_{\zeta_{ijkl}})^2 Var(\zeta_{tij1l})+(\lambda^{M}_{\zeta_{ijkl}})^2 Var(\zeta^{M}_{tjkl})}$
Occasion-specific method specificities		
Interchangeable	Rater	$OUMS(\pi_{rtij2l}) = \frac{(\lambda^{UM}_{\zeta_{ij2l}})^2 Var(\zeta^{UM}_{rtj2l})}{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})+(\lambda^{UM}_{\zeta_{ij2l}})^2 Var(\zeta^{UM}_{rtj2l})}$
Interchangeable	Rater	$OCMS(\pi_{rtij2l}) = \frac{(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})}{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})+(\lambda^{UM}_{\zeta_{ij2l}})^2 Var(\zeta^{UM}_{rtj2l})}$
Interchangeable	Target	$OCMS(\pi_{tij2l}) = \frac{(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})}{(\lambda_{\zeta_{ij2l}})^2 Var(\zeta_{tij1l})+(\lambda^{CM}_{\zeta_{ij2l}})^2 Var(\zeta^{CM}_{tj2l})}$
Structurally different	Target	$OMS(\pi_{tijkl}) = \frac{(\lambda^{M}_{\zeta_{ijkl}})^2 Var(\zeta^{M}_{tjkl})}{(\lambda_{\zeta_{ijkl}})^2 Var(\zeta_{tij1l})+(\lambda^{M}_{\zeta_{ijkl}})^2 Var(\zeta^{M}_{tjkl})}$

Note. Con: Consistency; CMS: Common Method Specificity; MS: Method Specificity; Spe: Specificity; UMS: Unique Method Specificity; O: Occasion; T: Trait.

4.9 Variance decompositions

Based on the definition of the LST-Com GRM, the latent response variables π_{tijkl} and π_{rtij2l} can be additively decomposed into different variance components. From Definition 4.1 and Theorem 4.1 it follows that the general measurement equations for the latent response variables in an LST-Com GRM of (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables are given by:

$$\pi_{tij1l} = \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\xi_{tij1} + \zeta_{tij1l} \tag{4.9.1}$$

$$\begin{aligned}\pi_{rtij2l} = {}& \alpha_{\xi ij2l} + \lambda_{\xi ij2l}\xi_{tij1} + \lambda^{CM}_{\xi ij2l}\xi^{CM}_{tij2} + \lambda_{\zeta ij2l}\zeta_{tij1l} + \lambda^{CM}_{\zeta ij2l}\zeta^{CM}_{tj2l} \\ & + \lambda^{UM}_{\xi ij2l}\xi^{UM}_{rtij2} + \lambda^{UM}_{\zeta ij2l}\zeta^{UM}_{rtj2l}\end{aligned} \tag{4.9.2}$$

$$\pi_{tijkl} = \alpha_{\xi ijkl} + \lambda_{\xi ijkl}\xi_{tij1} + \lambda^{M}_{\xi ijkl}\xi^{M}_{tijk} + \lambda_{\zeta ijkl}\zeta_{tij1l} + \lambda^{M}_{\zeta ijkl}\zeta^{M}_{tjkl} \qquad k > 2 \tag{4.9.3}$$

As the latent method variables are defined as latent residual variables, they are uncorrelated with their respective regressors. That is, due to the zero-covariances given in Equations (4.8.1) - (4.8.26), the different variance components can be separated. The variances of the latent response variables can therefore be additively decomposed as:

$$Var(\pi_{tij1l}) = (\lambda_{\xi ij1l})^2 Var(\xi_{tij1}) + Var(\zeta_{tij1l}) \tag{4.9.4}$$

$$\begin{aligned}Var(\pi_{rtij2l}) = {}& (\lambda_{\xi ij2l})^2 Var(\xi_{tij1}) + (\lambda^{CM}_{\xi ij2l})^2 Var(\xi^{CM}_{tij2}) + (\lambda_{\zeta ij2l})^2 Var(\zeta_{tij1l}) \\ & + (\lambda^{CM}_{\zeta ij2l})^2 Var(\zeta^{CM}_{tj2l}) + (\lambda^{UM}_{\xi ij2l})^2 Var(\xi^{UM}_{rtij2}) + (\lambda^{UM}_{\zeta ij2l})^2 Var(\zeta^{UM}_{rtj2l})\end{aligned} \tag{4.9.5}$$

$$\begin{aligned}Var(\pi_{tijkl}) = {}& (\lambda_{\xi ijkl})^2 Var(\xi_{tij1}) + (\lambda^{M}_{\xi ijkl})^2 Var(\xi^{M}_{tijk}) + (\lambda_{\zeta ijkl})^2 Var(\zeta_{tij1l}) \\ & + (\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl}) \qquad k > 2\end{aligned} \tag{4.9.6}$$

Then, analogous to the LST-Com model with continuous indicators, different variance coefficients can be defined. Definitions of the variance coefficients are given in Table 4.1. The variance coefficients can be meaningfully interpreted, as they are invariant under admissible transformations, as shown in Section 4.3.

Some of these variance coefficients correspond to the variance coefficients introduced in Koch et al. (2017), others are analogous to the coefficients introduced in Koch (2013, with the difference that they are defined based on the latent response variables π_{tijkl} and π_{rtij2l}). For interpretations of the coefficients see Section 4.1 or Koch et al. (2017). Note that definitions of reliability coefficients or the ICC are not included in Table 4.1, as they are identical to the definition of these coefficients in the LS-Com GRM (see Section 2.11).

4.10 Mean structure

The following theorem clarifies the mean structure of the latent variables in the LST-Com GRM. The mean structure is needed to derive the identification conditions in Section 4.11.

Theorem 4.7. ***(Mean Structure)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi, \boldsymbol{\lambda}^{\mathbf{M}}_\xi, \boldsymbol{\xi}^{\mathbf{UM}}_{rt}, \boldsymbol{\xi}^{\mathbf{CM}}_t, \boldsymbol{\xi}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ be an LST-Com GRM with (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables and conditional independence. Without loss of generality the first method (k=1) is chosen as reference method and the second method (k=2) as the set of interchangeable *methods. Then, for all $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$ it holds that*

$$\mathbb{E}(\xi^{UM}_{rtij2}) = 0 \tag{4.10.1}$$
$$\mathbb{E}(\xi^{CM}_{tij2}) = 0 \tag{4.10.2}$$
$$\mathbb{E}(\xi^{M}_{tijk}) = 0 \qquad k > 2 \tag{4.10.3}$$

$$\mathbb{E}(\zeta^{UM}_{rtj2l}) = 0 \tag{4.10.4}$$
$$\mathbb{E}(\zeta^{CM}_{tj2l}) = 0 \tag{4.10.5}$$
$$\mathbb{E}(\zeta^{M}_{tjkl}) = 0 \qquad k > 2 \tag{4.10.6}$$

$$\mathbb{E}(\zeta_{tij1l}) = 0 \tag{4.10.7}$$

$$\mathbb{E}(\pi_{tsijkl}) = \mathbb{E}(\pi_{tijkl}) - \kappa_{sijkl} \qquad k \neq 2 \tag{4.10.8}$$
$$\mathbb{E}(\pi_{rtsij2l}) = \mathbb{E}(\pi_{rtij2l}) - \kappa_{sij2l} \tag{4.10.9}$$

$$\mathbb{E}(\pi_{tij1l}) = \alpha_{\xi ij1l} + \lambda_{\xi ij1l}\, \mathbb{E}(\xi_{tij1}) \tag{4.10.10}$$
$$\mathbb{E}(\pi_{tijkl}) = \alpha_{\xi ijkl} + \lambda_{\xi ijkl}\, \mathbb{E}(\xi_{tij1}) \qquad k > 2 \tag{4.10.11}$$
$$\mathbb{E}(\pi_{rtij2l}) = \alpha_{\xi ij2l} + \lambda_{\xi ij2l}\, \mathbb{E}(\xi_{tij1}) \tag{4.10.12}$$

and in LST-Com GRMs defined with common latent trait factors ξ_{tj1}:

$$\mathbb{E}(\xi_{tij1}) = \delta_{ij1} + \lambda_{ij1}\, \mathbb{E}(\xi_{tj1}) \tag{4.10.13}$$

Proofs. *Mean Structure.*

Equations (4.10.1) - (4.10.7) *follow directly from the definition of the latent (trait and state residual) method variables as well as the latent state residual variables as residual variables in Definition 4.1 and the fact that residual variables have an expectation of zero (Steyer & Nagel, 2017, p. 323). Equations* (4.10.8) *and* (4.10.9) *follow directly from the definitions of the latent response variables π_{tsijkl} and $\pi_{rtsij2l}$ given in Definition 2.1. Equations* (4.10.10) - (4.10.12) *follow directly from Definition 4.1, Equations* (4.2.21) *and* (4.2.22) *in Theorem 4.1, as well as from Equations* (4.10.1) - (4.10.7)*. Equation* (4.10.13) *follows from Equation* (4.4.10)*. The proofs are straightforward and therefore left to the reader.*

Remarks. Equations (4.10.10) - (4.10.12) show that the expected value of the common latent response variables π_{tijkl}, $k \neq 2$, and π_{rtij2l} equal the expectation of the latent trait factors ξ_{tij1} if and only if $\alpha_{\xi ijkl} = 0$ and $\lambda_{\xi ijkl} = 1$. For models defined with common (non-indicator-specific) latent trait factors ξ_{tj1}, the expectation of the latent response variables π_{tijkl}, $k \neq 2$, and π_{rtij2l} equal the expectation of the latent trait factors ξ_{tj1} if and only if in addition $\delta_{ij1} = 0$ and $\lambda_{ij1} = 1$. The effect of different identification variants and parameter invariance settings on the interpretation of latent trait means and latent trait mean differences is discussed in Sections 4.11 and 4.12.

4.11 Identifiability

In order to assign a scale to each latent factor, either one factor loading per factor or the variance of the latent factor has to be fixed to a value larger than 0 (typically 1; Bollen, 1989).
As shown in Theorem 4.2, the latent response variables π_{rtijkl} and π_{tijkl}, their respective threshold variables κ_{sijkl}, and the variables α_{ijkl} are uniquely defined only up to translations. Consequently, the parameters $\alpha_{\xi ijkl}$ and κ_{sijkl} are not separately identifiable. The same holds for the parameters κ_{sijkl} and δ_{ij1} defined in Section 4.4 for the case of models with common latent trait factors over indicators. Theorem 4.8 defines identifiability conditions for the LST-Com GRM with indicator-specific latent trait variables ξ_{tij1}. Identification conditions for the model with common latent trait factors ξ_{tj1} can easily be derived from the conditions given in Theorem 4.8. Furthermore, recall that all latent method factors have an expectation of zero by definition.
In Equation (4.8.36) the total covariance matrix of the variables Y^*_{rtij2l} and Y^*_{tijkl} in an LST-Com GRM with strong conditional independence was represented as

$$\Sigma_T = \Lambda_{\xi B}\Phi_{\xi B}\Lambda'_{\xi B} + \Lambda_{\zeta B}\Phi_{\zeta B}\Lambda'_{\zeta B} + \Theta_B + \Lambda_{\xi W}\Phi_{\xi W}\Lambda'_{\xi W} + \Lambda_{\zeta W}\Phi_{\zeta W}\Lambda'_{\zeta W} + \Theta_W$$

where all non-zero elements in the residual variance-covariance matrices Θ_B and Θ_W are equal to 1. Theorem 4.8 then gives identification conditions for the LST-Com GRM parameters.

Theorem 4.8. ***(Identification of the LST-Com GRM)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi, \boldsymbol{\lambda}^{\mathbf{M}}_\xi, \boldsymbol{\xi}^{\mathbf{UM}}_{rt}, \boldsymbol{\xi}^{\mathbf{CM}}_t, \boldsymbol{\xi}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LST-Com GRM with* $(\xi_{tij1l}, \xi^{CM}_{tij2l}, \xi^{UM}_{rtij2l}, \xi^{M}_{tijkl}, \zeta^{CM}_{tij2l}, \zeta^{UM}_{tij2l}, \zeta^{M}_{tijkl})$*-congeneric variables and strong conditional independence as defined by Definitions 4.1, 4.2, and 4.3. The parameters of the LST-Com GRM with strong conditional independence and indicator-specific trait and state residual variables* $\xi_{tij1}, \xi^{CM}_{tij2}, \xi^{UM}_{rtij2}, \xi^{M}_{tijk}$, *and* ζ_{tij1l} *are identified if*

1. *either one factor loading* $\lambda_{\xi ij1l}, \lambda^{CM}_{\xi ij2l}, \lambda^{UM}_{\xi ij2l}, \lambda^{M}_{\xi ijkl}, \lambda_{\zeta ij1l}, \lambda^{CM}_{\zeta ij2l}, \lambda^{M}_{\zeta ijkl}$, *and* $\lambda^{UM}_{\zeta ij2l}$ *for each factor* $\xi_{tij1}, \xi^{CM}_{tij2}, \xi^{UM}_{rtij2}, \xi^{M}_{tijk}, \zeta_{tij1l}, \zeta^{CM}_{tj2l}, \zeta^{M}_{tjkl}$, *and* ζ^{UM}_{rtj2l}, *or the variance of the factors is set to any real value larger than 0, and*

2. *one of the following conditions hold:*

   (a) $i_j \geq 2$ *with* $i_j = 2$ *for all or for some* j, $j \geq 2$, $k \geq 2$, $l \geq 3$, *and* $\Phi_{\zeta B}$ *as well as* $\Phi_{\zeta W}$ *contain substantial (permissible) intercorrelations among the latent state residual variables as well as the latent state residual method variables,*

   (b) $i_j \geq 3$ *for all* j, $j \geq 1$, $k \geq 2$, $l \geq 3$, *and* $\Phi_{\zeta B}$ *contains substantial (permissible) intercorrelations between the latent state residual variables* ζ_{tij1l},

   (c) $i_j \geq 3$ *for all* j, $j \geq 1$, $k \geq 3$, *and* $l \geq 3$,

   and

3. $\alpha_{\xi ij1l}$ *is set to any real value (e.g., zero) for one reference-method indicator* Y_{tij1l} *per latent trait factor and either one threshold of the same indicator* Y_{tij1l} *or the mean of the latent trait factor* ξ_{tij1} *is set to any real value (e.g., zero) for all* i *and* j, *and*

4. *$\alpha_{\xi ijkl}$, $k \neq 1$, is set to any real value (e.g., zero) for all i, j, $k > 1$, for one measurement occasion l per latent trait factor, or one threshold of the same indicator Y_{tijkl} is fixed at any real value, and*

5. *one threshold per indicator is constrained to be invariant over measurement occasions for all i, j and k, that is $\kappa_{sijkl'} = \kappa_{sijkl}$ for a chosen value of s, and for all i, j, k and l, or the intercept $\alpha_{\xi ijkl}$ of the respective indicator is set to a any real value (e.g., zero).*

Remarks. Theorem 4.8 states the conditions under which the parameters of the LST-Com GRM are identified without further restrictions on loading, variance or threshold parameters than those that are necessary to assign a scale to the latent variables.

Condition (1) and (2) identify the parameters of the LST-Com GRM covariance structure (that is, the parameters in the matrices Λ_B, Λ_W, Φ_B, and Φ_W), given the polychoric correlations between the variables Y^*_{rtij2l} and Y^*_{tijkl} (or π_{rtij2l} and π_{tijkl}). These conditions also hold in the continuous-indicator LST-Com model and the identification of this part of the model was shown by Courvoisier (2006) and Koch (2013) and shall not be repeated here.

Note that Conditions 2 (a) - 2 (c) in Theorem 4.8 define the necessary numbers of indicators, constructs, methods and measurement occasions needed to identify the LST-Com GRM without imposing further assumptions. However, the LST-Com GRM can, obviously, also be estimated for smaller designs when a few additional assumptions are imposed. For instance, the LST-Com GRM with two indicators per construct, one construct, two methods, and three occasions of measurement is identified with the assumptions given in Theorem 4.8 when, additionally, the loading parameters of the method state residual variables, $\lambda^{CM}_{\zeta ij2l}$, $\lambda^{UM}_{\zeta ij2l}$, and $\lambda^{M}_{\zeta ijkl}$ are set to one. Similarly, the LST-Com GRM with three indicators per construct, two constructs, two methods, and two occasions of measurement is identified when, in addition to assumption 2 (b) in Theorem 4.8, the loading parameters $\lambda^{CM}_{\xi ij2l}$, $\lambda^{UM}_{\xi ij2l}$, and $\lambda^{M}_{\xi ijkl}$ are set to one and the loading parameters $\lambda_{\xi ij1l}$ are set invariant over measurement occasions for one reference-method indicator.

Conditions (3) - (5) are needed for the identification of the threshold variables κ_{sijkl}, intercept parameters $\alpha_{\xi ijkl}$ as well as of the means of the latent response variables μ_{ijkl} and latent trait variables. The identification of these parameters follows, with only minor differences, the same lines as in the LS-Com GRM and is explained in detail in Section 2.13.

Note that for interpretability reasons, it is the easiest to fix all intercepts, thresholds or means that are fixed for identification reasons to the value of zero. Furthermore, to enhance ease of interpretation, it is advisable to choose the identification variant that sets all intercept parameters $\alpha_{\xi ijkl}$ to zero. One exception might be the case, in which all threshold parameters are set invariant over measurement occasions for all categories s. In this case, the interpretation of $\alpha_{\xi ij1l}$ is straightforward, that is, it is the mean difference in the latent trait means of ξ_{tij1l} and ξ_{tij11}, while the distances between the thresholds of adjacent categories stay invariant over time (i.e., parallel shift in easiness for all categories of an item). For $k > 1$ and invariant threshold parameters κ_{sijkl} over time, $\alpha_{\xi ijkl}$ would represent a shift in the regression intercept of the conditional method bias for that item and method. See section 4.12 for more details on MI in the LST-Com GRM.

The LST-Com GRM with common latent trait factors for all indicators (i.e., ξ_{tj1}) needs similar identifiability conditions, with the addition that δ_{ij1} has to be set to any real value, preferably zero, for the same i and j as in Condition (3) of Theorem 4.8. Again, for the ease of interpretation, it is recommendable to set all δ_{ij1} to zero.

In the case were strong conditional independence does not hold, i.e., in the LST-Com GRM with conditional independence as defined by Definition 4.2 additional identification conditions are necessary.

As stated above, the correlations between ζ_{tij1l} and $\zeta_{tij1l'}$, ζ^{CM}_{tj2l} and $\zeta^{CM}_{tj2l'}$, ζ^{UM}_{rtj2l} and $\zeta^{UM}_{rtj2l'}$, as well as ζ^{M}_{tjkl} and $\zeta^{CM}_{tj2l'}$, for $l \neq l'$ are theoretically possible, it is, however, necessary to set some or all of these correlations to zero for identification reasons, depending on the respective model size. In the most typical case of an AR(1) process (i.e., only regressions between adjacent measurement occasions are included), the model is identifiable with at least three measurement occasions and some additional assumptions on the loading parameters, regression parameters and state residual variances (Cole et al., 2005). For a more detailed treatment of autoregressive processes in LST and LGC models see, e.g., Cole et al. (2005), Eid et al. (2012), the online appendices of Bishop et al. (2015), or Hamaker (2005), Hamaker et al. (2015). For a detailed illustration on how to apply these models in practice see Prenoveau (2016).

4.12 Measurement invariance over time

Definition 4.4. (LST-Com GRM with MI)

$\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_\xi, \boldsymbol{\lambda}_\xi, \boldsymbol{\xi}_t, \boldsymbol{\lambda}_\zeta, \boldsymbol{\zeta}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\xi, \boldsymbol{\lambda}^{\mathbf{CM}}_\xi, \boldsymbol{\lambda}^{\mathbf{M}}_\xi, \boldsymbol{\xi}^{\mathbf{UM}}_{rt}, \boldsymbol{\xi}^{\mathbf{CM}}_t, \boldsymbol{\xi}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{CM}}_\zeta, \boldsymbol{\lambda}^{\mathbf{M}}_\zeta, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ is called LST-Com GRM with (ξ_{tij1l}, ξ^{CM}_{tij2l}, ξ^{UM}_{rtij2l}, ξ^{M}_{tijkl}, ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables with measurement invariance if and only if Definition 4.1 and Theorem 4.1 apply, and for all $i \in I_j$, $j \in J$, $k \in K$, $s \in S_{ij}$ and for l, $l' \in L$ the following statements hold:

$$\kappa_{sijkl} = \kappa_{sijkl'} \tag{4.12.1}$$
$$\alpha_{\xi ijkl} = \alpha_{\xi ijkl'} \tag{4.12.2}$$
$$\lambda_{\xi ijkl} = \lambda_{\xi ijkl'} \tag{4.12.3}$$
$$\lambda_{\zeta ijkl} = \lambda_{\zeta ijkl'} \tag{4.12.4}$$
$$\lambda^{UM}_{\xi ij2l} = \lambda^{UM}_{\xi ij2l'} \tag{4.12.5}$$
$$\lambda^{CM}_{\xi ij2l} = \lambda^{CM}_{\xi ij2l'} \tag{4.12.6}$$
$$\lambda^{M}_{\xi ijkl} = \lambda^{M}_{\xi ijkl'} \qquad k > 2 \tag{4.12.7}$$
$$\lambda^{UM}_{\zeta ij2l} = \lambda^{UM}_{\zeta ij2l'} \tag{4.12.8}$$
$$\lambda^{CM}_{\zeta ij2l} = \lambda^{CM}_{\zeta ij2l'} \tag{4.12.9}$$
$$\lambda^{M}_{\zeta ijkl} = \lambda^{M}_{\zeta ijkl'} \qquad k > 2 \tag{4.12.10}$$

Remarks. To obtain a pure state-variability model, strong measurement invariance as defined above (Definition 4.4) has to be established. In accordance with Geiser, Keller, et al. (2015) and Koch et al. (2017) it is recommended to establish at least strong MI when applying the LST-Com GRM model to real data. LST models with non-invariant intercepts may confound measurement non-invariance with true trait change (Geiser, Keller, et al., 2015). Similarly, changes in the loadings can reflect either trait change or measurement bias (changes in item discrimination), which are not distinguishable in LST models (Geiser, Keller, et al., 2015). Furthermore, Geiser, Keller, et al. (2015) found LST models with non-invariant trait loadings and intercepts to fit data adequately even in the presence of inter-individual differences in trait change (i.e., slope variance > 0 in LGC models). In the case of medium to large slope variances present in LGC models, the incorrect use of an LST model can lead to biased consistency and occasion-specificity coefficients (Geiser, Keller, et al., 2015). Hence, if trait change or measurement non-invariance of the trait loadings and intercepts are found in LST models, it is recommendable to test whether inter-individual differences in trait change exist by applying, e.g., a growth curve model (as presented in Section 5). In the case that inter-individual differences in intraindividual trait change are assumed between a priori specified periods of time (including more than one measurement occasion each), an unambiguous separation of trait change and measurement invariance is possible by applying double-trait models, as for instance shown in Eid and Hoffmann (1998) or Steyer et al. (2015).

4.13 The LST-Com GRM in LST-R theory

Recently, Steyer et al. (2015) proposed a revision of LST theory (LST-R) that explicitly takes into account that persons might change over the course of time, defining trait change on the basis of a different conceptualization of the random experiment. In their definition of the random experiment, the probability space Ω explicitly includes persons' experiences between measurement occasions and defines occasion-specific person projections that take into account these experiences as part of a time-specific person variable (person-at-time-l). This definition implies a temporal ordering of the elements in the probability space. For a minimal mono-method design (e.g., including only the targets' self-reports) and three measurement occasions the elements ω of Ω would have the following structure (Steyer et al., 2015),

$$\omega = (u_0, e_1, s_1, o_1, e_2, s_2, o_2, e_3, s_3, o_3), \tag{4.13.1}$$

where u_0 is the person at time 0 (the sampling time point), e_1 denotes the experiences of the person after time 0 and before the first measurement occasion (the assessment at time 1), s_1 is the situation at the assessment on measurement occasion 1, and o_1 are the observables on measurement occasion 1. Then u_1, the person at time 1, is given by (u_0, e_1), while u_2, the person at time 2, is given by $(u_0, e_1, s_1, o_1, e_2)$. Hence, the person variable at time $l > 1$ is defined as $U_l : \Omega \to \Omega_0 \times \Omega_{E_1} \times \Omega_{S_1} \times \Omega_{O_1} \times \ldots \times \Omega_0 \times \Omega_{E_{l-1}} \times \Omega_{S_{l-1}} \times \Omega_{O_{l-1}} \times \Omega_{E_l}$. The situation variable at time l is defined as $S_l := \Omega \to \Omega_{S_l}$ in LST-R theory. The trait variables are then defined as the expectations of the conditional distribution of an observed variable Y_{il} given the person at time l, i.e., $\xi_{il} := \mathbb{E}[Y_{il} \mid U_l]$, while the state residuals are defined as residuals with respect to this expectation, i.e., $\zeta_{il} := \mathbb{E}[Y_{il} \mid U_l, S_l] - \mathbb{E}[Y_{il} \mid U_l]$ (Steyer et al., 2015). This definition explicitly takes into account that persons can change due to experiences and that there is no person without a past.

Koch et al. (2017) show how LST-R theory can be combined with multi-method modeling approaches for interchangeable and structurally different methods for the continuous-indicator LST-Com model. They clarify which modifications have to be made in order to make the LST-Com model compatible with the revised version of LST theory. The same modifications hold for the LST-Com GRM, too.

It is noteworthy that, although LST and LST-R theory diverge in their definition of the random experiment and the probability space, the vast majority of models that can be specified on the basis of these theories are identical. That is, most empirical applications of LST models will lead to identical results, whether theoretically building on LST or LST-R theory.

While state residual variables defined based on LST-R theory are uncorrelated over time by definition, Eid, Holtmann, Santangelo, and Ebner-Priemer (in press) have recently shown that LST models with autoregressive effects can nevertheless be formulated in the framework of LST-R theory.

Chapter 5

Latent Growth Curve (LGC-Com) Graded Response Model

5.1 Introduction to the LGC-Com GRM

This chapter introduces a longitudinal multilevel MTMM Latent Growth-Curve graded response model for measurement designs combining structurally different and interchangeable methods (LGC-Com GRM). The model is based on the definition of the LST-Com GRM given in Section 4.2 as well as the definition of the random experiment and latent response variables in Section 2.4. Note that the LGC-Com GRM as defined in the following differs from the continuous-indicator LGC-Com model as defined by Koch (2013), as growth is not only modeled for the reference method but differential change can also occur in the non-reference methods.

Just as latent change models (Geiser et al., 2010; McArdle & Hamagami, 2001; Steyer et al., 1997, 2000), latent growth curve (LGC) models (Bollen & Curran, 2006; McArdle & Epstein, 1987; McArdle & Nesselroade, 2003) allow to model inter-individual differences in intra-individual change. However, in contrast to LC models, LGC models aim at modeling change as a (linear or non-linear) function of time.

The LGC model defined in the following builds on LST theory, that is, it combines features of LGC models with the distinction of trait (change) and state variability processes. Different hybrid models, combining features of both state variability processes (LST models) and growth curve models, have, for instance, been proposed by McArdle (1988), Tisak and Tisak (2000), Eid et al. (2012), and Bishop et al. (2015). Models that allow for this combination are models including multiple indicators per measurement occasion (so-called second-order LGC models; Geiser, Keller, & Lockhart, 2013; Leite, 2007). Geiser et al. (2013) showed how to define the latent variables in LGC models on the basis of LST theory, demonstrating that second-order LGC models represent a restrictive variant of LST change models.

Besides the possibility to separate true trait change processes from occasion-specific variability and measurement error (Sayer & Cumsille, 2001), multiple-indicator LGC models bear additional advantages over single-indicator LGC models, such as yielding more accurate reliability estimates (Geiser et al., 2013), providing a greater power to detect individual differences in change (von Oertzen, Hertzog, Lindenberger, & Ghisletta, 2010), a greater flexibility in modeling complex change patterns (Mayer, Geiser, Infurna, & Fiege, 2013), and the possibility to model indicator-specific patterns of trait change (Bishop et al., 2015).

The reference-method part of the model corresponds to the indicator-specific growth model (ISGM) as introduced by Bishop et al. (2015), which allows for indicator-specific growth processes. In a first step, the latent trait variables ξ_{tijkl} belonging to a measurement occasion $l > 1$ are decomposed into an initial trait variable ξ_{tijk1} and a latent change variable $(\xi_{tijkl} - \xi_{tijk1})$,

$$\xi_{tijkl} = \xi_{tijk1} + (\xi_{tijkl} - \xi_{tijk1}) \tag{5.1.1}$$

assuming that the change between different measurement occasions follows a specific function, e.g., a linear function,

$$(\xi_{tijkl} - \xi_{tijk1}) = (l-1)(\xi_{tijk2} - \xi_{tijk1}) \tag{5.1.2}$$

with

$$\mathcal{I}_{tijk} = \xi_{tijk1} \tag{5.1.3}$$
$$\mathcal{S}_{tijk} = (\xi_{tijk2} - \xi_{tijk1}) \tag{5.1.4}$$

where $\mathcal{I}_{tijk}$ is called the intercept factor and $\mathcal{S}_{tijk}$ the slope factor. The assumption of a linear growth trajectory could be replaced and extended to model non-linear change trajectories, by additionally adding factors for quadratic change, i.e., $(l-1)^2$, or cubic change, i.e., $(l-1)^3$.
Having defined intercept and slope factors for both reference and non-reference method indicators, these can be regressed on each other, again following the CTC(M-1) approach for multimethod data (Eid, 2000; Eid et al., 2003, 2008), with

$$\mathcal{I}^{CM}_{tij2} = \mathcal{I}_{tij2} - \mathbb{E}[\mathcal{I}_{tij2} \mid \mathcal{I}_{tij1}] \tag{5.1.5}$$
$$\mathcal{I}^{M}_{tijk} = \mathcal{I}_{tijk} - \mathbb{E}[\mathcal{I}_{tijk} \mid \mathcal{I}_{tij1}] \qquad k > 2 \tag{5.1.6}$$
$$\mathcal{S}^{CM}_{tij2} = \mathcal{S}_{tij2} - \mathbb{E}[\mathcal{S}_{tij2} \mid \mathcal{S}_{tij1}] \tag{5.1.7}$$
$$\mathcal{S}^{M}_{tijk} = \mathcal{S}_{tijk} - \mathbb{E}[\mathcal{S}_{tijk} \mid \mathcal{S}_{tij1}] \qquad k > 2 \tag{5.1.8}$$

where the dependence of the non-reference method intercept / slope variables $\mathcal{I}_{tijk}$ / $\mathcal{S}_{tijk}$ on the reference method intercept / slope variable $\mathcal{I}_{tij1}$ / $\mathcal{S}_{tij1}$ of the same indicator i and construct j can be described by linear transformations:

$$\mathbb{E}[\mathcal{I}_{tijk} \mid \mathcal{I}_{tij1}] = \alpha_{\mathcal{I}ijk} + \lambda_{\mathcal{I}ijk}\mathcal{I}_{tij1} \tag{5.1.9}$$

and

$$\mathbb{E}[\mathcal{S}_{tijk} \mid \mathcal{S}_{tij1}] = \alpha_{\mathcal{S}ijk} + \lambda_{\mathcal{S}ijk}\mathcal{S}_{tij1} \tag{5.1.10}$$

The latent intercept (common) method variables $\mathcal{I}^{CM}_{tij2}$ and $\mathcal{I}^{M}_{tijk}$, $k > 2$, represent that part of the latent trait variables of the non-reference methods on the first measurement occasion that cannot be explained by the reference-method traits on the first measurement occasion. The latent slope (common) method variables $\mathcal{S}^{CM}_{tij2}$ and $\mathcal{S}^{M}_{tijk}$, $k > 2$, represent that part of the latent growth in the non-reference method traits that cannot be explained by the growth in the reference method traits. For instance, a value on the latent slope method variable $\mathcal{S}^{M}_{tijk}$ indicates to which degree a structurally different rater over- or underestimates the linear slope of the growth trajectory for the respective target with respect to the value that is expected based on the targets' self-reported linear growth trajectory. Hence, the latent slope method variables $\mathcal{S}^{M}_{tijk}$ represent the part of the latent growth in a construct as rated by the structurally different non-reference method rater that cannot be explained by the self-reported latent growth.
In a similar logic as used for the latent traits ξ_{tijkl}, the latent unique method trait variables ξ^{UM}_{rtij2l} belonging to a measurement occasion $l > 1$ can be decomposed into an initial unique method trait variable ξ^{UM}_{rtij21} and a latent unique method change variable $(\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21})$,

$$\xi^{UM}_{rtij2l} = \xi^{UM}_{rtij21} + (\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21}) \tag{5.1.11}$$

with

$$(\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21}) = (l-1)(\xi^{UM}_{rtij22} - \xi^{UM}_{rtij21}) \tag{5.1.12}$$

Again, intercept and slope factors are defined as:

$$\mathcal{I}^{UM}_{rtij2} = \xi^{UM}_{rtij21} \tag{5.1.13}$$

$$\mathcal{S}^{UM}_{rtij2} = (\xi^{UM}_{rtij22} - \xi^{UM}_{rtij21}) \tag{5.1.14}$$

The latent unique method intercept variables $\mathcal{I}^{UM}_{rtij2}$ represent the unique method trait variables on the first occasion of measurement, while the unique method slope variables $\mathcal{S}^{UM}_{rtij2}$ represent the growth (linear change) in the unique method trait variables, i.e., the change in the rater-specific view between adjacent measurement occasions, that is not shared with the other raters or the target and not influenced by the specific situations. Note that, as the latent unique method trait variables are defined as residual variables, their expectation is zero by definition. This implies that the intercept and slope factors $\mathcal{I}^{UM}_{rtij2}$ and $\mathcal{S}^{UM}_{rtij2}$ also have a mean of zero by definition and that there can be no mean change, that is, no linear average change or trend in the unique method trait variables. The unique method slope variables represent change in the relative position of an individual rater's rating to the expected rating over all raters per target. Hence, the unique method slope variance indicates to which degree there is a change in the individual raters' views relative to the other interchangeable raters' views between adjacent measurement occasions.

As in latent change models, an important prerequisite for the application of LGC models is strong measurement invariance across time (Ferrer, Balluerka, & Widaman, 2008). Only if strong MI holds, the growth components can be meaningfully interpreted, as change is investigated with respect to the same latent variables. In the present LGC model definition, intercept and slope factors are defined separately for the reference method and the non-reference method trait factors [see Equations (5.1.1)-(5.1.4)], before regressing non-reference method intercept / slope variables on reference method intercept / slope variables. Also, intercept and slope factors are defined based on the indicator- and occasion-specific latent trait factors ξ_{tijkl}. Therefore, there are no loading parameters that need to be invariant over time. However, MI over time of the threshold parameters is required not only for the reference method indicators but also for the non-reference method indicators.

An LGC-Com GRM with indicator-specific latent (method) intercept and slope factors is depicted in Figure 5.1, an LGC-Com GRM with common latent (method) intercept and slope factors is depicted in Figure 5.2.

Formal definitions and detailed explanations of the latent variables, as well as theorems on their uniqueness, admissible transformations and meaningful statements are presented in Sections 5.1 - 5.4. Furthermore, independence assumptions imposed on the LGC-Com GRM, a detailed variance decomposition and the identifiability of the model are derived in Sections 5.7 - 5.8. Note that measurement invariance will not be separately defined, as measurement invariance of the threshold variables is a prerequisite for meaningful interpretations of the presented LGC model, while loading parameters on the trait-level of the LGC-Com GRM are invariant by definition. For details on measurement invariance for the occasion-specific part of the LGC-Com GRM see the respective definition for the LST-Com GRM (Section 4.12).

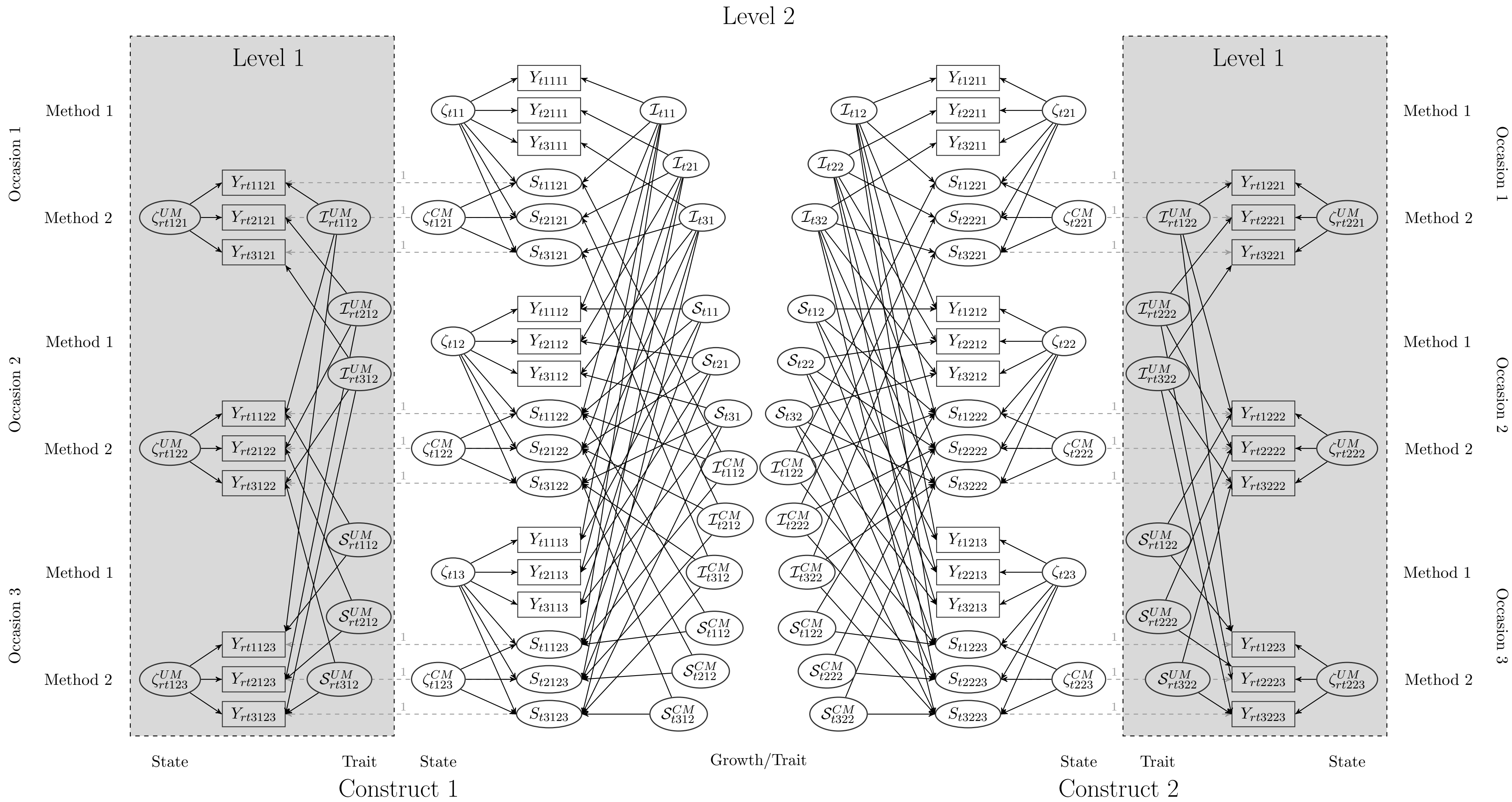

Figure 5.1: Path diagram of the Latent-Growth-Curve-Com graded response model with indicator-specific latent intercept and slope variables $\mathcal{I}_{tij}$, $\mathcal{S}_{tij}$, $\mathcal{I}^{CM}_{tij2}$, $\mathcal{S}^{CM}_{tij2}$, $\mathcal{I}^{UM}_{rtij2}$ and $\mathcal{S}^{UM}_{rtij2}$ and common latent state residual variables ζ_{tjl}. The model is depicted for one structurally different method and one set of interchangeable methods on three measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For the sake of clarity, correlations between latent variables and loading parameters are omitted. Note that loading parameters of the latent intercept and slope variables are restricted in order to model a linear growth trajectory. Correlations that are not permissible in the depicted LGC-Com GRM are all correlations between any (method) intercept variable $\mathcal{I}$ or (method) slope variable $\mathcal{S}$ and any state residual (method) variable ζ, correlations between the latent intercept and the latent intercept (common) method variables of the same construct j and indicator i, correlations between the latent slope and the latent slope (common) method variables of the same construct j and indicator i, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method; M: method; S: latent state variable; $\mathcal{S}$: latent slope variable; UM: unique method; $\mathcal{I}$: latent intercept variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

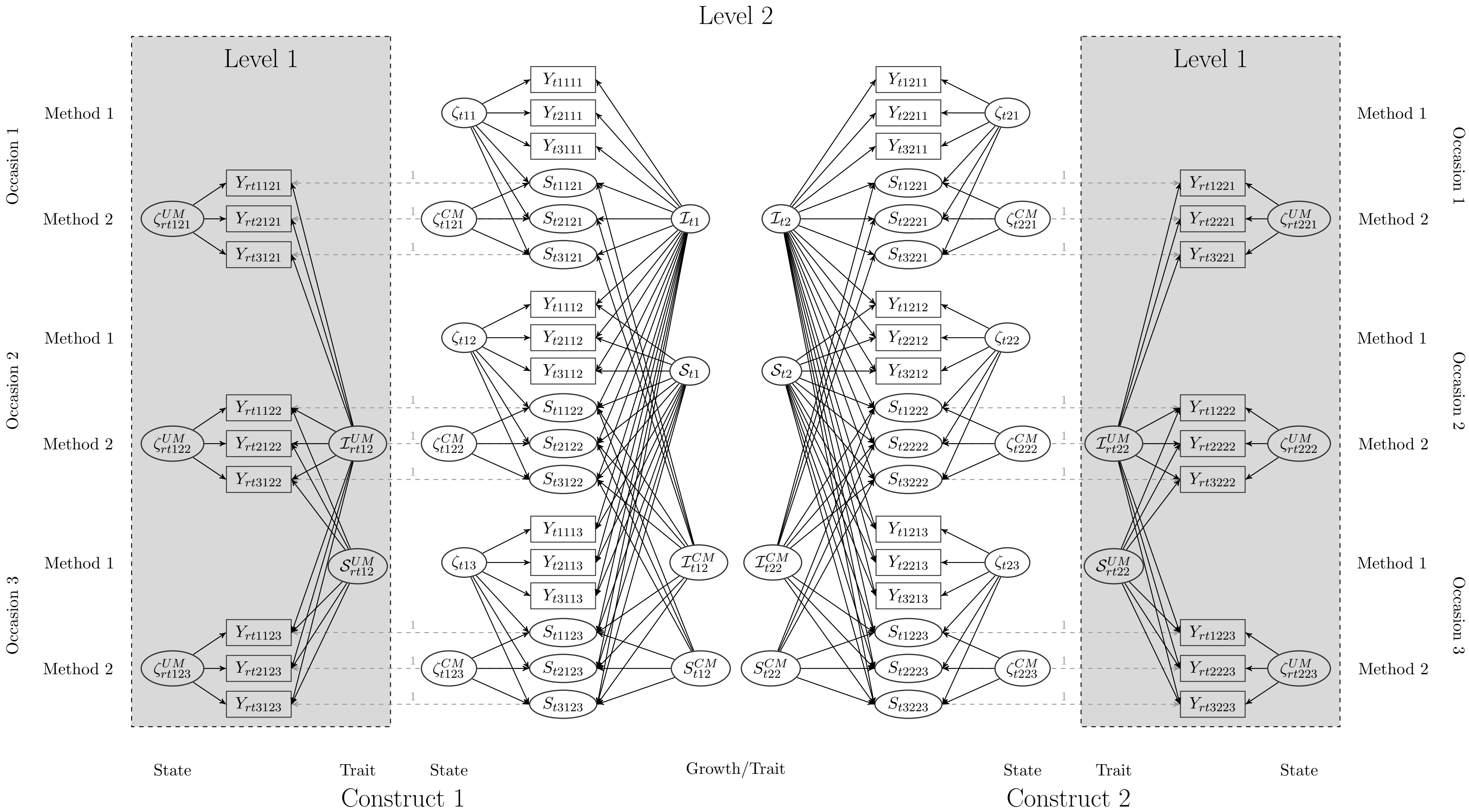

Figure 5.2: Path diagram of the Latent-Growth-Curve-Com graded response model with common latent intercept and slope variables $\mathcal{I}_{tj}$, $\mathcal{S}_{tj}$, $\mathcal{I}^{CM}_{tj2}$, $\mathcal{S}^{CM}_{tj2}$, $\mathcal{I}^{UM}_{rtj2}$ and $\mathcal{S}^{UM}_{rtj2}$ and latent state residual variables ζ_{tjl}. The model is depicted for one structurally different method and one set of interchangeable methods on three measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are, however, not linearly linked but probabilistically linked to the latent variables by a probit link. For the sake of clarity, correlations between latent variables and loading parameters are omitted. Note that loading parameters of the latent intercept and slope variables are restricted in order to model a linear growth trajectory. Correlations that are not permissible in the depicted LGC-Com GRM are all correlations between any (method) intercept variable $\mathcal{I}$ or (method) slope variable $\mathcal{S}$ and any state residual (method) variable ζ, correlations between the latent intercept and the latent intercept (common) method variables of the same construct j, correlations between the latent slope and the latent slope (common) method variables of the same construct j, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method; M: method; S: latent state variable; $\mathcal{S}$: latent slope variable; UM: unique method; $\mathcal{I}$: latent intercept variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

5.2 Formal Definition of the LGC-Com GRM

In the following the LGC-Com GRM is formally defined building on the definition of the LST-Com GRM in section 4.2. Note that the following definition of the LGC-Com GRM is not analogous to the continuous-indicator LGC-Com model as defined by Koch (2013), but differs from it in that growth is not only assumed for the reference method but differential change can also occur in the non-reference methods.

Definition 5.1. (LGC-Com GRM)

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ *and* $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ *on a probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *are variables of an LGC-Com graded response model if Conditions* (a) *to* (e) *in Definition 2.1, Conditions (a), (c), and (f) in Definition 4.1 and the following conditions hold:*

(a) For all $i \in I_j$, $j \in J$, $k \in K$, *and* $l, l' \in L$, $l \neq l'$, *it holds that*

$$\kappa_{sijkl} = \kappa_{sijkl'} \tag{5.2.1}$$

(b) Then, without loss of generality, the latent trait variables ξ_{tijkl} *belonging to a measurement occasion* $l > 1$ *can be decomposed into an initial trait variable* ξ_{tijk1} *and a latent change variable* $(\xi_{tijkl} - \xi_{tijk1})$*:*

$$\xi_{tijkl} = \xi_{tijk1} + (\xi_{tijkl} - \xi_{tijk1}) \tag{5.2.2}$$

(c) For each indicator i, construct j, method k and measurement occasion $l > 1$, *it holds that*

$$(\xi_{tijkl} - \xi_{tijk1}) = (l-1)(\xi_{tijk2} - \xi_{tijk1}) \tag{5.2.3}$$

Define the intercept factors $\mathcal{I}_{tijk}$ *and slope factors* $\mathcal{S}_{tijk}$ *by:*

$$\mathcal{I}_{tijk} := \xi_{tijk1} \tag{5.2.4}$$
$$\mathcal{S}_{tijk} := (\xi_{tijk2} - \xi_{tijk1}) \tag{5.2.5}$$

(d) The following latent variables are random variables on $(\Omega, \mathcal{A}, \mathcal{P})$ *with finite first- and second-order moments:*

$$\mathcal{I}^{CM}_{tij2} = \mathcal{I}_{tij2} - \mathbb{E}[\mathcal{I}_{tij2} \mid \mathcal{I}_{tij1}] \tag{5.2.6}$$
$$\mathcal{I}^{M}_{tijk} = \mathcal{I}_{tijk} - \mathbb{E}[\mathcal{I}_{tijk} \mid \mathcal{I}_{tij1}] \qquad k > 2 \tag{5.2.7}$$
$$\mathcal{S}^{CM}_{tij2} = \mathcal{S}_{tij2} - \mathbb{E}[\mathcal{S}_{tij2} \mid \mathcal{S}_{tij1}] \tag{5.2.8}$$
$$\mathcal{S}^{M}_{tijk} = \mathcal{S}_{tijk} - \mathbb{E}[\mathcal{S}_{tijk} \mid \mathcal{S}_{tij1}] \qquad k > 2 \tag{5.2.9}$$

(e) *The latent unique method trait variables ξ^{UM}_{rtij2l} belonging to a measurement occasion $l > 1$ can be decomposed into an initial unique method trait variable ξ^{UM}_{rtij21} and a latent unique method change variable $(\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21})$:*

$$\xi^{UM}_{rtij2l} = \xi^{UM}_{rtij21} + (\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21}) \tag{5.2.10}$$

Then, for each indicator i, construct j, and measurement occasion $l > 1$, it holds that,

$$(\xi^{UM}_{rtij2l} - \xi^{UM}_{rtij21}) = (l-1)(\xi^{UM}_{rtij22} - \xi^{UM}_{rtij21}) \tag{5.2.11}$$

Define the latent unique method intercept factors $\mathcal{I}^{UM}_{rtij2}$ and unique method slope factors $\mathcal{S}^{UM}_{rtij2}$ by:

$$\mathcal{I}^{UM}_{rtij2} := \xi^{UM}_{rtij21} \tag{5.2.12}$$
$$\mathcal{S}^{UM}_{rtij2} := (\xi^{UM}_{rtij22} - \xi^{UM}_{rtij21}) \tag{5.2.13}$$

(f) *For each indicator i of construct j measured by a non-reference method $(k \neq 1)$, there are constants $\alpha_{\mathcal{I}ijk} \in \mathbb{R}$ and $\lambda_{\mathcal{I}ijk} \in \mathbb{R}^+$ such that*

$$\mathbb{E}[\mathcal{I}_{tijk} \mid \mathcal{I}_{tij1}] = \alpha_{\mathcal{I}ijk} + \lambda_{\mathcal{I}ijk}\mathcal{I}_{tij1} \tag{5.2.14}$$

(g) *For each indicator i of construct j measured by a non-reference method $(k \neq 1)$, there are constants $\alpha_{\mathcal{S}ijk} \in \mathbb{R}$ and $\lambda_{\mathcal{S}ijk} \in \mathbb{R}^+$ such that*

$$\mathbb{E}[\mathcal{S}_{tijk} \mid \mathcal{S}_{tij1}] = \alpha_{\mathcal{S}ijk} + \lambda_{\mathcal{S}ijk}\mathcal{S}_{tij1} \tag{5.2.15}$$

Remarks. The preceding definition of the LGC-Com GRM is based on the definition of the LST-Com GRM, replacing the assumption of perfectly correlated latent trait variables ξ_{tij1l} over measurement occasion by the assumption of linearity in presumed trait change between measurement occasions. Note that the assumption of a linear growth trajectory made in Equation (5.2.3) of Definition 5.1 (c) could be adapted and extended to model non-linear change trajectories, such as quadratic change (i.e., $(l-1)^2$) or cubic change (i.e., $(l-1)^3$). Assumption (a) of Definition 5.1 defines strong measurement invariance for the latent trait variables (as given by Equations 4.2.6 and 4.2.12), necessary to ensure that the variables ξ_{tijkl} measure the same latent construct over time and the change variables $(\xi_{tijkl} - \xi_{tijk1})$ can be meaningfully interpreted.

Assumption (d) defines the latent intercept (common) method variables $\mathcal{I}^{CM}_{tij2}$ and $\mathcal{I}^{M}_{tijk}$, $k > 2$, as well as the latent slope (common) method variables $\mathcal{S}^{CM}_{tij2}$ and $\mathcal{S}^{M}_{tijk}$, $k > 2$. The latent intercept (common) method variables $\mathcal{I}^{CM}_{tij2}$ and $\mathcal{I}^{M}_{tijk}$, $k > 2$, represent that part of the latent trait variables of the non-reference methods on the first measurement occasion that cannot be explained by the reference-method traits on the first measurement occasion. The latent slope (common) method variables $\mathcal{S}^{CM}_{tij2}$ and $\mathcal{S}^{M}_{tijk}$, $k > 2$, represent that part of the latent growth in the non-reference method traits that cannot be explained by the growth in the reference method traits.

Assumption (e) defines the latent unique method intercept and slope variables $\mathcal{I}^{UM}_{rtij2}$ and $\mathcal{S}^{UM}_{rtij2}$. The latent unique method intercept variables $\mathcal{I}^{UM}_{rtij2}$ represent the unique method trait variables on the first

measurement occasion, while the unique method slope variables $\mathcal{S}^{UM}_{rtij2}$ represent the growth (linear change) in the unique method trait variables, i.e., the change in the rater-specific view between adjacent measurement occasions, that is not shared with the other raters or the target and not influenced by the specific situations. Note that, as the latent unique method trait variables are defined as residual variables, their expectation is zero by definition. This implies, that there can be no mean change, that is, no linear average change or trend in the unique method trait variables. The unique method slope variables represent change in the relative position of an individual rater's rating to the expected rating over all raters per target. Hence, the unique method slope variance indicates to which degree there is a change in the individual raters' views relative to the other interchangeable raters' views between adjacent measurement occasions.
Equations (5.2.14) and (5.2.15) state the assumptions that the dependence of the non-reference method intercept / slope variables $\mathcal{I}_{tijk}$ / $\mathcal{S}_{tijk}$ on the reference method intercept / slope variable $\mathcal{I}_{tij1}$ / $\mathcal{S}_{tij1}$ of the same indicator i and construct j can be described by linear transformations. Note that the linearity of the dependence of the non-reference method intercept variables $\mathcal{I}_{tijk}$ on the reference method slope variable $\mathcal{I}_{tij1}$ of the same indicator i and construct j is only repeated here and was already stated by Equation (4.2.12) of Definition 4.1 (b), as the latent intercept variables are defined as $\mathcal{I}_{tijk} = \xi_{tijk1}$.
As the LGC-Com GRM model is derived from the LST-Com GRM, all psychometric statements with respect to existence, uniqueness, admissible transformations or meaningfulness of the latent state residual (method) variables correspond to those of the LST-Com GRM. That is, (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congenerity also holds in the LGC-Com GRM, as stated in the following theorem. Psychometric properties of the latent trait (that is, intercept and slope) variables in the LGC-Com GRM are derived in the following sections.

Theorem 5.1. ***(Existence)***

The random variables $\{Y_{rt1111}, \ldots, Y_{rtijkl}, \ldots, Y_{rtc_d def}\}$ and $\{Y_{t1111}, \ldots, Y_{tijkl}, \ldots, Y_{tc_d def}\}$ are (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables of an LGC-Com GRM if and only if the conditions in Definition 5.1 hold. Then, for each $r \in R$, $t \in T$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, there are real-valued random variables ζ^{CM}_{tj2l}, ζ^{UM}_{tj2l}, and ζ^{M}_{tjkl} on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and constants ($\lambda^{CM}_{\zeta ij2l}$, $\lambda^{UM}_{\zeta ij2l}$, $\lambda^{M}_{\zeta ijkl}$) $\in \mathbb{R}^+$ such that:

$$\zeta^{UM}_{rtij2l} = \lambda^{UM}_{\zeta ij2l} \zeta^{UM}_{rtj2l} \tag{5.2.16}$$
$$\zeta^{CM}_{tij2l} = \lambda^{CM}_{\zeta ij2l} \zeta^{CM}_{tj2l} \tag{5.2.17}$$
$$\zeta^{M}_{tijkl} = \lambda^{M}_{\zeta ijkl} \zeta^{M}_{tjkl} \qquad \forall \; k > 2 \tag{5.2.18}$$

5.3 Uniqueness, admissible transformations and meaningful statements

Uniqueness and admissible transformations of the latent state residual (method) variables correspond to those in the LST-Com GRM, as the definition of these variables is identical in the two models. However, as the remaining latent variables differ from those of the LST-Com GRM, for completeness, a theorem on the uniqueness of the latent variables in the LGC-Com GRM is given in the following. Note that measurement invariance of the threshold parameters is assumed, i.e., $\kappa_{sijkl} = \kappa_{sijkl'} \; \forall k$.

Theorem 5.2. ***(Admissible transformations and uniqueness)***

1. Admissible Transformations
Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_t$, $\boldsymbol{\lambda}_\zeta$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{CM}}_\zeta$, $\boldsymbol{\lambda}^{\mathbf{M}}_\zeta$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LGC-Com GRM with:*

$$\boldsymbol{\pi}_{rt} = (\pi_{rt1121}, \ldots, \pi_{rtij2l}, \ldots, \pi_{rtc_d d2f})^T \tag{5.3.1}$$
$$\boldsymbol{\pi}_t = (\pi_{t1111}, \ldots, \pi_{tijkl}, \ldots, \pi_{tc_d def})^T \qquad k \neq 2 \tag{5.3.2}$$
$$\boldsymbol{\kappa} = (\kappa_{11111}, \ldots, \kappa_{sijkl}, \ldots, \kappa_{(q_{c_d d}-1)c_d def})^T \tag{5.3.3}$$

$$\boldsymbol{\alpha}_{\mathcal{I}} = (\alpha_{\mathcal{I}112}, \ldots, \alpha_{\mathcal{I}ijk}, \ldots, \alpha_{\mathcal{I}c_d de})^T \tag{5.3.4}$$
$$\boldsymbol{\lambda}_{\mathcal{I}} = (\lambda_{\mathcal{I}112}, \ldots, \lambda_{\mathcal{I}ijk}, \ldots, \lambda_{\mathcal{I}c_d de})^T \tag{5.3.5}$$
$$\boldsymbol{\mathcal{I}}_t = (\mathcal{I}_{t111}, \ldots, \mathcal{I}_{tij1}, \ldots, \mathcal{I}_{tc_d d1})^T \tag{5.3.6}$$
$$\boldsymbol{\alpha}_{\mathcal{S}} = (\alpha_{\mathcal{S}112}, \ldots, \alpha_{\mathcal{S}ijk}, \ldots, \alpha_{\mathcal{S}c_d de})^T \tag{5.3.7}$$
$$\boldsymbol{\lambda}_{\mathcal{S}} = (\lambda_{\mathcal{S}112}, \ldots, \lambda_{\mathcal{S}ijk}, \ldots, \lambda_{\mathcal{S}c_d de})^T \tag{5.3.8}$$
$$\boldsymbol{\mathcal{S}}_t = (\mathcal{S}_{t111}, \ldots, \mathcal{S}_{tij1}, \ldots, \mathcal{S}_{tc_d d1})^T \tag{5.3.9}$$
$$\boldsymbol{\lambda}_\zeta = (\lambda_{\zeta 1121}, \ldots, \lambda_{\zeta ijkl}, \ldots, \lambda_{\zeta c_d def})^T \qquad k > 1 \tag{5.3.10}$$
$$\boldsymbol{\zeta}_t = (\zeta_{t1111}, \ldots, \zeta_{tij1l}, \ldots, \zeta_{tc_d d1f})^T \tag{5.3.11}$$

$$\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt} = (\mathcal{I}^{\mathbf{UM}}_{rt112}, \ldots, \mathcal{I}^{\mathbf{UM}}_{rtij2}, \ldots, \mathcal{I}^{\mathbf{UM}}_{rtc_d d2})^T \tag{5.3.12}$$
$$\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t = (\mathcal{I}^{\mathbf{CM}}_{t112}, \ldots, \mathcal{I}^{\mathbf{CM}}_{tij2}, \ldots, \mathcal{I}^{\mathbf{CM}}_{tc_d d2})^T \tag{5.3.13}$$
$$\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t = (\mathcal{I}^{\mathbf{M}}_{t113}, \ldots, \mathcal{I}^{\mathbf{M}}_{tijk}, \ldots, \mathcal{I}^{\mathbf{M}}_{tc_d de})^T \qquad k > 2 \tag{5.3.14}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt} = (\mathcal{S}^{\mathbf{UM}}_{rt112}, \ldots, \mathcal{S}^{\mathbf{UM}}_{rtij2}, \ldots, \mathcal{S}^{\mathbf{UM}}_{rtc_d d2})^T \tag{5.3.15}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t = (\mathcal{S}^{\mathbf{CM}}_{t112}, \ldots, \mathcal{S}^{\mathbf{CM}}_{tij2}, \ldots, \mathcal{S}^{\mathbf{CM}}_{tc_d d2})^T \tag{5.3.16}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t = (\mathcal{S}^{\mathbf{M}}_{t113}, \ldots, \mathcal{S}^{\mathbf{M}}_{tijk}, \ldots, \mathcal{S}^{\mathbf{M}}_{tc_d de})^T \qquad k > 2 \tag{5.3.17}$$

$$\boldsymbol{\lambda}^{\mathbf{UM}}_\zeta = (\lambda^{\mathbf{UM}}_{\zeta 1121}, \ldots, \lambda^{\mathbf{UM}}_{\zeta ij2l}, \ldots, \lambda^{\mathbf{UM}}_{\zeta c_d d2f})^T \tag{5.3.18}$$
$$\boldsymbol{\lambda}^{\mathbf{CM}}_\zeta = (\lambda^{\mathbf{CM}}_{\zeta 1121}, \ldots, \lambda^{\mathbf{CM}}_{\zeta ij2l}, \ldots, \lambda^{\mathbf{CM}}_{\zeta c_d d2f})^T \tag{5.3.19}$$
$$\boldsymbol{\lambda}^{\mathbf{M}}_\zeta = (\lambda^{\mathbf{M}}_{\zeta 1131}, \ldots, \lambda^{\mathbf{M}}_{\zeta ijkl}, \ldots, \lambda^{\mathbf{M}}_{\zeta c_d def})^T \qquad k > 2 \tag{5.3.20}$$
$$\boldsymbol{\zeta}^{\mathbf{UM}}_{rt} = (\zeta^{\mathbf{UM}}_{rt121}, \ldots, \zeta^{\mathbf{UM}}_{rtj2l}, \ldots, \zeta^{\mathbf{UM}}_{rtd2f})^T \tag{5.3.21}$$
$$\boldsymbol{\zeta}^{\mathbf{CM}}_t = (\zeta^{\mathbf{CM}}_{t121}, \ldots, \zeta^{\mathbf{CM}}_{tj2l}, \ldots, \zeta^{\mathbf{CM}}_{td2f})^T \tag{5.3.22}$$
$$\boldsymbol{\zeta}^{\mathbf{M}}_t = (\zeta^{\mathbf{M}}_{t131}, \ldots, \zeta^{\mathbf{M}}_{tjkl}, \ldots, \zeta^{\mathbf{M}}_{tdef})^T \qquad k > 2 \tag{5.3.23}$$

If for all $r \in R$, $t \in T$, $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$:

$$\pi'_{tijkl} = \pi_{tijkl} + \nu_{ijk} \qquad k \neq 2 \tag{5.3.24}$$
$$\pi'_{rtij2l} = \pi_{rtij2l} + \nu_{ij2} \tag{5.3.25}$$
$$\kappa'_{sijkl} = \kappa_{sijkl} + \nu_{ijk} \tag{5.3.26}$$

$$\alpha'_{\mathcal{I}ijk} = \alpha_{\mathcal{I}ijk} + \nu_{ijk} - \lambda_{\mathcal{I}ijk}\nu_{ij1} \qquad k > 1 \tag{5.3.27}$$

$$\zeta'^{\mathbf{UM}}_{rtj2l} = \beta^{\mathbf{UM}}_{\zeta j2l}\zeta^{\mathbf{UM}}_{rtj2l} \tag{5.3.28}$$
$$\zeta'^{\mathbf{CM}}_{tj2l} = \beta^{\mathbf{CM}}_{\zeta j2l}\zeta^{\mathbf{CM}}_{tj2l} \tag{5.3.29}$$
$$\zeta'^{\mathbf{M}}_{tjkl} = \beta^{\mathbf{M}}_{\zeta jkl}\zeta^{\mathbf{M}}_{tjkl} \qquad k > 2 \tag{5.3.30}$$
$$\lambda'^{\mathbf{UM}}_{\zeta ij2l} = \lambda^{\mathbf{UM}}_{\zeta ij2l}/\beta^{\mathbf{UM}}_{\zeta j2l} \tag{5.3.31}$$
$$\lambda'^{\mathbf{CM}}_{\zeta ij2l} = \lambda^{\mathbf{CM}}_{\zeta ij2l}/\beta^{\mathbf{CM}}_{\zeta j2l} \tag{5.3.32}$$
$$\lambda'^{\mathbf{M}}_{\zeta ijkl} = \lambda^{\mathbf{M}}_{\zeta ijkl}/\beta^{\mathbf{M}}_{\zeta jkl} \qquad k > 2 \tag{5.3.33}$$

where $\beta^{\mathbf{UM}}_{\zeta j2l}$, $\beta^{\mathbf{CM}}_{\zeta j2l}$, $\beta^{\mathbf{M}}_{\zeta jkl} \in \mathbb{R}_+$, *and* $\gamma_{\mathcal{I}ij1}$, $\gamma_{\mathcal{S}ij1}$, $\nu_{ijkl} \in \mathbb{R}$.
Then $\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}'_{rt}$, $\boldsymbol{\pi}'_t$, $\boldsymbol{\kappa}'$, $\boldsymbol{\alpha}'_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}'_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}'^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}'^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}'^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}'^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}'^{\mathbf{M}}_t\rangle$ *is an LGC-Com GRM, too, with*

$$\boldsymbol{\pi}'_{rt} = (\pi'_{rt1121}, \ldots, \pi'_{rtij2l}, \ldots, \pi'_{rtc_d d2f})^T \tag{5.3.34}$$
$$\boldsymbol{\pi}'_t = (\pi'_{t1111}, \ldots, \pi'_{tijkl}, \ldots, \pi'_{tc_d def})^T \qquad k \neq 2 \tag{5.3.35}$$
$$\boldsymbol{\kappa}' = (\kappa'_{11111}, \ldots, \kappa'_{sijkl}, \ldots, \kappa'_{(q_{c_d d}-1)c_d def})^T \tag{5.3.36}$$

$$\boldsymbol{\alpha}'_{\mathcal{I}} = (\alpha'_{\mathcal{I}112}, \ldots, \alpha'_{\mathcal{I}ijk}, \ldots, \alpha'_{\mathcal{I}c_d de})^T \tag{5.3.37}$$
$$\boldsymbol{\lambda}_{\mathcal{I}} = (\lambda_{\mathcal{I}112}, \ldots, \lambda_{\mathcal{I}ijk}, \ldots, \lambda_{\mathcal{I}c_d de})^T \tag{5.3.38}$$
$$\boldsymbol{\mathcal{I}}'_t = (\mathcal{I}'_{t111}, \ldots, \mathcal{I}'_{tij1}, \ldots, \mathcal{I}'_{tc_d d1})^T \tag{5.3.39}$$
$$\boldsymbol{\alpha}_{\mathcal{S}} = (\alpha_{\mathcal{S}112}, \ldots, \alpha_{\mathcal{S}ijk}, \ldots, \alpha_{\mathcal{S}c_d de})^T \tag{5.3.40}$$
$$\boldsymbol{\lambda}_{\mathcal{S}} = (\lambda_{\mathcal{S}112}, \ldots, \lambda_{\mathcal{S}ijk}, \ldots, \lambda_{\mathcal{S}c_d de})^T \tag{5.3.41}$$
$$\boldsymbol{\mathcal{S}}_t = (\mathcal{S}_{t111}, \ldots, \mathcal{S}_{tij1}, \ldots, \mathcal{S}_{tc_d d1})^T \tag{5.3.42}$$
$$\boldsymbol{\lambda}_{\zeta} = (\lambda_{\zeta 1121}, \ldots, \lambda_{\zeta ijkl}, \ldots, \lambda_{\zeta c_d def})^T \qquad k > 1 \tag{5.3.43}$$
$$\boldsymbol{\zeta}_t = (\zeta_{t1111}, \ldots, \zeta_{tij1l}, \ldots, \zeta_{tc_d d1f})^T \tag{5.3.44}$$

$$\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_t = (\mathcal{I}^{\mathbf{UM}}_{t112}, \ldots, \mathcal{I}^{\mathbf{UM}}_{tij2}, \ldots, \mathcal{I}^{\mathbf{UM}}_{tc_d d2})^T \tag{5.3.45}$$
$$\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t = (\mathcal{I}^{\mathbf{CM}}_{t112}, \ldots, \mathcal{I}^{\mathbf{CM}}_{tij2}, \ldots, \mathcal{I}^{\mathbf{CM}}_{tc_d d2})^T \tag{5.3.46}$$
$$\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t = (\mathcal{I}^{\mathbf{M}}_{t113}, \ldots, \mathcal{I}^{\mathbf{M}}_{tijk}, \ldots, \mathcal{I}^{\mathbf{M}}_{tc_d de})^T \qquad k > 2 \tag{5.3.47}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_t = (\mathcal{S}^{\mathbf{UM}}_{t112}, \ldots, \mathcal{S}^{\mathbf{UM}}_{tij2}, \ldots, \mathcal{S}^{\mathbf{UM}}_{tc_d d2})^T \qquad k > 2 \tag{5.3.48}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t = (\mathcal{S}^{\mathbf{CM}}_{t112}, \ldots, \mathcal{S}^{\mathbf{CM}}_{tij2}, \ldots, \mathcal{S}^{\mathbf{CM}}_{tc_d d2})^T \tag{5.3.49}$$
$$\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t = (\mathcal{S}^{\mathbf{M}}_{t113}, \ldots, \mathcal{S}^{\mathbf{M}}_{tijk}, \ldots, \mathcal{S}^{\mathbf{M}}_{tc_d de})^T \qquad k > 2 \tag{5.3.50}$$

$$\boldsymbol{\lambda}'^{\mathbf{UM}}_{\zeta} = (\lambda'^{\mathbf{UM}}_{\zeta 1121}, \ldots, \lambda'^{\mathbf{UM}}_{\zeta ij2l}, \ldots, \lambda'^{\mathbf{UM}}_{\zeta c_d 2f})^T \tag{5.3.51}$$
$$\boldsymbol{\lambda}'^{\mathbf{CM}}_{\zeta} = (\lambda'^{\mathbf{CM}}_{\zeta 1121}, \ldots, \lambda'^{\mathbf{CM}}_{\zeta ij2l}, \ldots, \lambda'^{\mathbf{CM}}_{\zeta c_d 2f})^T \tag{5.3.52}$$
$$\boldsymbol{\lambda}'^{\mathbf{M}}_{\zeta} = (\lambda'^{\mathbf{M}}_{\zeta 1131}, \ldots, \lambda'^{\mathbf{M}}_{\zeta ijkl}, \ldots, \lambda'^{\mathbf{M}}_{\zeta c_d def})^T \qquad k > 2 \tag{5.3.53}$$
$$\boldsymbol{\zeta}'^{\mathbf{UM}}_{rt} = (\zeta'^{\mathbf{UM}}_{rt121}, \ldots, \zeta'^{\mathbf{UM}}_{rtj2l}, \ldots, \zeta'^{\mathbf{UM}}_{rtd2f})^T \tag{5.3.54}$$
$$\boldsymbol{\zeta}'^{\mathbf{CM}}_{t} = (\zeta'^{\mathbf{CM}}_{t121}, \ldots, \zeta'^{\mathbf{CM}}_{tj2l}, \ldots, \zeta'^{\mathbf{CM}}_{td2f})^T \tag{5.3.55}$$
$$\boldsymbol{\zeta}'^{\mathbf{M}}_{t} = (\zeta'^{\mathbf{M}}_{t131}, \ldots, \zeta'^{\mathbf{M}}_{tjkl}, \ldots, \zeta'^{\mathbf{M}}_{tdef})^T \qquad k > 2 \tag{5.3.56}$$

2. Uniqueness
If both $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_{\mathcal{I}}, \boldsymbol{\lambda}_{\mathcal{I}}, \boldsymbol{\mathcal{I}}_t, \boldsymbol{\alpha}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}, \boldsymbol{\mathcal{S}}_t, \boldsymbol{\lambda}_{\zeta}, \boldsymbol{\zeta}_t, \boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}, \boldsymbol{\mathcal{I}}^{\mathbf{CM}}_{t}, \boldsymbol{\mathcal{I}}^{\mathbf{M}}_{t}, \boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt}, \boldsymbol{\mathcal{S}}^{\mathbf{CM}}_{t}, \boldsymbol{\mathcal{S}}^{\mathbf{M}}_{t}, \boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}, \boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}, \boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_{t}, \boldsymbol{\zeta}^{\mathbf{M}}_{t}\rangle$ *and*
$\mathcal{M}' = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}'_{rt}, \boldsymbol{\pi}'_t, \boldsymbol{\kappa}', \boldsymbol{\alpha}'_{\mathcal{I}}, \boldsymbol{\lambda}_{\mathcal{I}}, \boldsymbol{\mathcal{I}}'_t, \boldsymbol{\alpha}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}, \boldsymbol{\mathcal{S}}_t, \boldsymbol{\lambda}_{\zeta}, \boldsymbol{\zeta}_t, \boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}, \boldsymbol{\mathcal{I}}^{\mathbf{CM}}_{t}, \boldsymbol{\mathcal{I}}^{\mathbf{M}}_{t}, \boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{t}, \boldsymbol{\mathcal{S}}^{\mathbf{CM}}_{t}, \boldsymbol{\mathcal{S}}^{\mathbf{M}}_{t}, \boldsymbol{\lambda}'^{\mathbf{UM}}_{\zeta}, \boldsymbol{\lambda}'^{\mathbf{CM}}_{\zeta}, \boldsymbol{\lambda}'^{\mathbf{M}}_{\zeta}, \boldsymbol{\zeta}'^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}'^{\mathbf{CM}}_{t}, \boldsymbol{\zeta}'^{\mathbf{M}}_{t}\rangle$ *are LGC-Com GRMs, then for each* $i \in I_j$, $j \in J$, $k \in K$, *and* $l \in L$ *there are* $\gamma_{\mathcal{I}ij1}$, $\gamma_{\mathcal{S}ij1}$, $\nu_{ijkl} \in \mathbb{R}$, *and* $\beta^{\mathbf{UM}}_{\zeta j2l}$, $\beta^{\mathbf{CM}}_{\zeta j2l}$, $\beta^{\mathbf{M}}_{\zeta jkl} \in \mathbb{R}^+$, *such that Equations* (5.3.24) *to* (5.3.33) *hold.*

Remarks. As stated in Theorem 5.2, the uniqueness and admissible transformations for the latent state residual (method) variables (i.e., ζ_{tij1l}, $\zeta^{\mathbf{UM}}_{rtj2l}$, $\zeta^{\mathbf{CM}}_{tj2l}$, $\zeta^{\mathbf{M}}_{tjkl}$) and their parameters (i.e., $\lambda_{\zeta ijkl}$, $\lambda^{\mathbf{UM}}_{\zeta ij2l}$, $\lambda^{\mathbf{CM}}_{\zeta ij2l}$, $\lambda^{\mathbf{M}}_{\zeta ijkl}$), as well as the latent response variables π_{tijkl} and π_{rtij2l} in the LGC-Com GRM are identical to those in the LST-Com GRM. That is, the latent state residual variables ζ_{tij1l} as well as the loading parameters $\lambda_{\zeta ijkl}$ are uniquely defined in the LGC-Com GRM with indicator-specific state residual variables ζ_{tij1l}. Again, this is the case as any translation of the latent response variables π_{tijkl} directly translates to the same translation for the latent trait variables $\xi_{tijkl} = \mathbb{E}[S_{tijkl} \mid p_T]$, with $S_{tijkl} = \pi_{tijkl}$, $k \neq 2$, and $S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$, and thereby does not affect their residuals $\zeta_{tijkl} = S_{tijkl} - \xi_{tijkl}$.
As in the LST-Com GRM, the common state residual method variables $\zeta^{\mathbf{CM}}_{tj2l}$, $\zeta^{\mathbf{UM}}_{rtj2l}$, and $\zeta^{\mathbf{M}}_{tjkl}$ and their corresponding loading parameters are uniquely defined only up to similarity transformations. The parameters π'_{rtij2l}, π'_{tijkl}, and κ_{sijkl} are uniquely defined up to translations by a constant.
Theorem 5.2 reveals that the intercept variables $\mathcal{I}_{tij1}$ are uniquely defined only up to translations. Although this translation is not explicitly stated in Theorem 5.2, it follows directly from the definition of the latent intercept variables and the fact that the latent response variables π'_{rtij2l} and π'_{tijkl} are only uniquely defined up to translations. That is, the latent intercept variables $\mathcal{I}_{tijk}$ are defined as $\mathcal{I}_{tijk} := \xi_{tijk1}$ with $\xi_{tijk1} = \mathbb{E}[S_{tijk1} \mid p_T]$ and $S_{tijk1} = \pi_{tijk1}$, $k \neq 2$, and $S_{tij21} = \mathbb{E}[\pi_{rtij21} \mid p_T, p_{TS_l}]$, such that any translation of π_{tijk1} by ν_{tijk} directly leads to a translation of $\mathcal{I}_{tijk}$ by the same amount ν_{tijk}.
In contrast, the latent slope variables $\mathcal{S}_{tij1}$ are uniquely defined in the LGC-Com GRM. This is the case as measurement invariance of the threshold parameters is a prerequisite for the LGC-Com GRM, such that the translation of the threshold and latent response variables is time invariant, too, which is why the parameter ν_{ijk} does not have an index l. Consequently, the latent slope variables $\mathcal{S}_{tijk}$, which are defined as $\mathcal{S}_{tijk} := \xi_{tijk2} - \xi_{tijk1}$ are not affected by the translations of π_{tijk2} and π_{tijk1} by ν_{ijk} as $\mathcal{S}_{tijk} = \xi_{tijk2} - \xi_{tijk1} = (\xi_{tijk2} + \nu_{ijk}) - (\xi_{tijk1} - \nu_{ijk})$.
The same holds for the parameters $\alpha_{\mathcal{I}ijk}$ and $\alpha_{\mathcal{S}ijk}$, that is, the parameters $\alpha_{\mathcal{I}ijk}$ are uniquely defined only up to translations while the parameters $\alpha_{\mathcal{S}ijk}$ are uniquely defined.
The loading parameters $\lambda_{\mathcal{I}ijk}$ and $\lambda_{\mathcal{S}ijk}$ are uniquely defined in the LGC-Com GRM. Furthermore, the

intercept and slope unique method, common method and method variables $\mathcal{I}^{\mathbf{UM}}_{rtij2}$, $\mathcal{I}^{\mathbf{CM}}_{tij2}$, $\mathcal{I}^{\mathbf{M}}_{tijk}$, $\mathcal{S}^{\mathbf{UM}}_{rtij2}$, $\mathcal{S}^{\mathbf{CM}}_{tij2}$, and $\mathcal{S}^{\mathbf{M}}_{tijk}$ are uniquely defined in the LGC-Com GRM. Again, this is the case as they are defined as zero-mean residual variables, which are not affected by the translation of regressand and regressor. That is, there are no admissible transformations (except for the identity transformation) for the variables $\mathcal{S}_{tijk}$, $\alpha_{\mathcal{S}ijk}$, $\mathcal{I}^{\mathbf{UM}}_{rtij2}$, $\mathcal{I}^{\mathbf{CM}}_{tij2}$, $\mathcal{I}^{\mathbf{M}}_{tijk}$, $\mathcal{S}^{\mathbf{UM}}_{rtij2}$, $\mathcal{S}^{\mathbf{CM}}_{tij2}$, $\mathcal{S}^{\mathbf{M}}_{tijk}$, $\lambda_{\mathcal{I}ijk}$, $\lambda_{\mathcal{S}ijk}$, ζ_{tij1l}, and $\lambda_{\zeta ijkl}$, and meaningful statements can directly be made about their absolute values.
To see that the latent intercept $\mathcal{I}_{tij1}$ are only uniquely defined up to translations, let π'_{tijkl} and κ'_{sijkl} be defined as given by Equations (5.3.24) and (5.3.26). Then, it holds that, for instance for $k = 1$ and $l = 1$,

$$\begin{aligned}\pi_{tsij11} &= \pi_{tij11} - \kappa_{tij11}\\ &= \xi_{tij11} + \zeta_{tij11} - \kappa_{tij11}\\ &= \mathcal{I}_{tij1} + \zeta_{tij11} - \kappa_{tij11}\\ &= \mathcal{I}_{tij1} + \nu_{ij11} + \zeta_{tij11} - \kappa_{tij11} - \nu_{ij11}\\ &= (\mathcal{I}_{tij1} + \nu_{ij11} + \zeta_{tij11}) - (\kappa_{tij11} + \nu_{ij11})\\ &= \pi'_{tij11} - \kappa'_{tij11}\end{aligned}$$

As the latent intercept variables $\mathcal{I}_{tij1}$ and their parameters $\alpha_{\mathcal{I}ijk}$ are uniquely defined only up to translations, they are measured on a difference scale. Therefore, meaningful statements regarding the intercept variables $\mathcal{I}_{tij1}$ are statements on their differences: for ω_1, $\omega_2 \in \Omega$, $r \in R$, $t \in T$, $i \in I_j$, $j \in J$, $k \in K$, and $l \in L$, it holds that

$$\mathcal{I}_{tij1}(\omega_1) - \mathcal{I}_{tij1}(\omega_2) = \mathcal{I}'_{tij1}(\omega_1) - \mathcal{I}'_{tij1}(\omega_2)$$

as

$$\begin{aligned}\mathcal{I}'_{tij1}(\omega_1) - \mathcal{I}'_{tij1}(\omega_2) &= \big(\mathcal{I}_{tij1}(\omega_1) + \gamma_{\mathcal{I}ij1}\big) - \big(\mathcal{I}_{tij1}(\omega_2) + \gamma_{\mathcal{I}ij1}\big)\\ &= \mathcal{I}_{ij1}(\omega_1) - \mathcal{I}_{ij1}(\omega_2)\end{aligned}$$

It follows that statements about differences in the latent intercept variables between different constructs are also only meaningful for the differences between persons.
For meaningful statements regarding the remaining variables and their parameters see Section 4.3, i.e., meaningful statements in the LST-Com GRM.

5.4 Common latent intercept, slope, and residual state factors

The LGC-Com GRM defined in Section 5.2 could also be defined with common latent state residual factors ζ_{tj1l} for all indicators belonging to the same construct j and measurement occasion l, instead of the indicator-specific latent state residual variables ζ_{tij1l}. The definition of common latent state residual factors is identical as in the LST-Com GRM and is described in detail in Section 4.4.
Additionally, it could be assumed, that the latent intercept variables $\mathcal{I}_{tij1}$ and $\mathcal{I}_{ti'j1}$ of different indicators $i, i' \in I_j$, $i \neq i'$, belonging to the same construct j are linear transformations of each other (and hence perfectly correlated). That is, it can be assumed that for each construct j, measured by the reference method $(k = 1)$, and for each pair $(i, i') \in I_j \times I_j$, $(i \neq i')$, there are constants $\lambda_{\mathcal{I}ii'j1} \in \mathbb{R}^+$ and $\delta_{\mathcal{I}ii'j1} \in \mathbb{R}$ such that

$$\mathcal{I}_{tij1} = \delta_{\mathcal{I}ii'j1} + \lambda_{\mathcal{I}ii'j1}\mathcal{I}_{ti'j1}. \tag{5.4.1}$$

Again, this implies the existence of common latent intercept variables $\mathcal{I}_{tj1}$ and constants $\lambda_{\mathcal{I}ij1} \in \mathbb{R}^+$ and $\delta_{\mathcal{I}ij1} \in \mathbb{R}$ such that

$$\mathcal{I}_{tij1} = \delta_{\mathcal{I}ij1} + \lambda_{\mathcal{I}ij1}\mathcal{I}_{tj1}. \tag{5.4.2}$$

From Equation (5.4.2) it is obvious that the common latent intercept variables $\mathcal{I}_{tj1}$ are only uniquely defined up to linear transformations, while the coefficients $\lambda_{\mathcal{I}ij1}$ are uniquely defined up to similarity transformations and the coefficients $\delta_{\mathcal{I}ij1}$ up to translations by a real constant. Note that the coefficients $\lambda_{\mathcal{I}ijk}$ and $\alpha_{\mathcal{I}ijk}$ in the Equation for $\mathbb{E}[\mathcal{I}_{tijk} \mid \mathcal{I}_{tij1}]$ (see Equation 5.2.14) change from the model with indicator-specific latent intercept variables $\mathcal{I}_{tij1}$ to the model with common latent intercept variables $\mathcal{I}_{tj1}$, as they are now the coefficients of the regression of the non-reference method intercept variables $\mathcal{I}_{tijk}$ on $\mathcal{I}_{tj1}$ (instead of on $\mathcal{I}_{tij1}$).

Analogous assumptions of perfectly correlated latent slope variables $\mathcal{S}_{tij1}$ and $\mathcal{S}_{ti'j1}$ of different indicators $i, i' \in I_j$, $i \neq i'$, belonging to the same construct j can be made, defining common latent slope variables $\mathcal{S}_{tj1}$. The same holds for the latent intercept and slope method variables $\mathcal{I}^{CM}_{tij2}$, $\mathcal{I}^{M}_{tijk}$, $\mathcal{I}^{UM}_{rtij2}$, $\mathcal{S}^{CM}_{tij2}$, $\mathcal{S}^{M}_{tijk}$, and $\mathcal{S}^{UM}_{rtij2}$. The definitions of the common latent slope variables $\mathcal{S}_{tj1}$ and conclusions of their definition are perfectly analogous to the case of the common latent intercept variables described above. In the case of the method intercept or slope variables, the common latent variables are uniquely defined only up to similarity transformations, as their expectations are zero by definition (compare definition of common latent method trait variables in the LST-Com GRM, Section 4.3 and 4.4).

As they are uniquely defined up to linear transformations, meaningful statements regarding the common latent intercept factors $\mathcal{I}_{tj1}$ are statements on the ratio of differences between different values of $\mathcal{I}_{tj1}$, that is, for ω_1, ω_2, ω_3, $\omega_4 \in \Omega$, $t \in T$, and $j \in J$ it holds that

$$\frac{\mathcal{I}_{tj1}(\omega_1) - \mathcal{I}_{tj1}(\omega_2)}{\mathcal{I}_{tj1}(\omega_3) - \mathcal{I}_{tj1}(\omega_4)} = \frac{\mathcal{I}'_{tj1}(\omega_1) - \mathcal{I}'_{tj1}(\omega_2)}{\mathcal{I}'_{tj1}(\omega_3) - \mathcal{I}'_{tj1}(\omega_4)} \tag{5.4.3}$$

For the new loading parameters $\lambda_{\mathcal{I}ijk}$ in an LGC-Com GRM with common latent intercept factors $\mathcal{I}_{tj1}$ meaningful statements are statements regarding the ratio of $\lambda_{\mathcal{I}ijk}$ and $\lambda_{\mathcal{I}i'jk'}$ for $i, i' \in I_j$, $i = i'$ or $\neq i'$, $k, k' \in K$, $k = k'$ or $k \neq k'$, as

$$\frac{\lambda_{\mathcal{I}ijk}}{\lambda_{\mathcal{I}i'jk'}} = \frac{\lambda'_{\mathcal{I}ijk}}{\lambda'_{\mathcal{I}i'jk'}} \tag{5.4.4}$$

This result is a direct consequence of the definition of the variables analogous to the results in Section 4.3 and the proof is left to the reader. This result on admissible transformations of the common latent intercept factors and their loading parameters applies in analogous manner to the common latent slope variables and their respective loading parameters. Admissible transformations for the common latent method intercept and slope parameters are the multiplications with positive real numbers, as they are measured on a ratio scale. Meaningful statements for these parameters are statements regarding the ratio of specific values of the factor loadings or the ratio of the values of the latent trait intercept or slope method factors (see Section 4.3).

Note that, as π_{tijkl} and κ_{sijkl} are only uniquely defined up to translations, the coefficients $\delta_{\mathcal{I}ij1}$, $\delta_{\mathcal{S}ij1}$, $\alpha_{\mathcal{I}ijk}$, $\alpha_{\mathcal{S}ijk}$ and all of the coefficients κ_{sij1l} for the same indicator i and construct j are not separately identifiable. For further restrictions imposed on the mean structure, as well as the coefficients $\alpha_{\mathcal{I}ijk}$, $\alpha_{\mathcal{S}ijk}$, $\delta_{\mathcal{I}ij1}$, $\delta_{\mathcal{S}ij1}$ and κ_{sijkl} due to identifiability considerations refer to Sections 5.9 and 5.10.

An LGC-Com GRM with common latent intercept, slope, method intercept and method slope variables is depicted in Figure 5.2.

5.5 True score variables

The definition of latent true score variables for ordered categorical variables in the LGC-Com GRM is identical to that of the LS-Com GRM and was given in Definition 2.2 in section 2.7.

5.6 Factor analytical representation

The LGC-Com GRM presented above can also be represented as a factor model for ordinal data. As this representation does not depend on the specific model, the factor-analytical representation of the LGC-Com GRM is identical to that of the LS-Com GRM as defined in section 2.8.

5.7 Independence assumptions and testability

5.7.1 LGC-Com GRM with conditional independence

In order to derive testable consequences of the LGC-Com GRM, several independence assumptions have to be introduced. As the LGC-Com GRM is defined on the basis of the LST-Com GRM, these independence assumptions correspond to those made in the LST-Com GRM.

Definition 5.2. (LGC-Com GRM with conditional independence)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\mathcal{I}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\mathcal{S}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\mathcal{I}_{rt}^{\mathbf{UM}}$, $\mathcal{I}_t^{\mathbf{CM}}$, $\mathcal{I}_t^{\mathbf{M}}$, $\mathcal{S}_{rt}^{\mathbf{UM}}$, $\mathcal{S}_t^{\mathbf{CM}}$, $\mathcal{S}_t^{\mathbf{M}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{UM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{CM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{M}}$, $\boldsymbol{\zeta}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\zeta}_t^{\mathbf{CM}}$, $\boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ be an LGC-Com GRM with (ζ_{tij2l}^{CM}, ζ_{tij2l}^{UM}, ζ_{tijkl}^{M})-congeneric variables. $\mathcal{M}$ is called LGC-Com GRM with conditional independence if and only if the assumptions given in Definition 4.2 hold.

Remarks. The independence assumptions given in Definitions 2.3 and 4.2 have the same meaning in the LGC-Com GRM as they have in the LST-Com GRM and are explained in detail in the remarks to Definition 4.2. Furthermore, all of the conditional independence assumptions given in Definition 4.2 (also see Definition 2.3) imply consequences regarding the conditional and unconditional distributions of the observed variables Y_{tijkl} and Y_{rtij2l} as well as a specific covariance structure of the latent variables π_{tijkl} and π_{rtij2l} in the LGC-Com GRM. The conditional independence assumptions given in Definition 4.2 (see Definition 2.3) impose testable consequences on the covariance structure of the LGC-Com GRM. These are, with minor differences for newly defined latent intercept and slope variables, analogous to those of the LST-Com GRM. They are stated in Section 5.7.3 and the following theorem.

Theorem 5.3. ***(LGC-Com GRM with conditional independence)***

$\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\mathcal{I}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\mathcal{S}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\mathcal{I}_{rt}^{\mathbf{UM}}$, $\mathcal{I}_t^{\mathbf{CM}}$, $\mathcal{I}_t^{\mathbf{M}}$, $\mathcal{S}_{rt}^{\mathbf{UM}}$, $\mathcal{S}_t^{\mathbf{CM}}$, $\mathcal{S}_t^{\mathbf{M}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{UM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{CM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{M}}$, $\boldsymbol{\zeta}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\zeta}_t^{\mathbf{CM}}$, $\boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ be an LGC-Com GRM with conditional independence. Then,

for all $i, i' \in I_j$, $j, j' \in J$, $k \in K$, $l, l' \in L$, *and* y_{rtij2l}, $y_{tijkl} \in S_{ij}$ *it holds that:*

$$P\left(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}(Y_{rtij2l}=y_{rtij2l})\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\leq e,k\neq 2}(Y_{tijkl}=y_{tijkl})\;\middle|\;\pi_{t1111},\ldots,\pi_{tc_ddef},\pi_{rt1121},\ldots,\pi_{rtc_dd2f}\right) \tag{5.7.1}$$
$$=\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}\prod_{k\leq e,k\neq 2}P(Y_{tijkl}=y_{tijkl}\mid\pi_{tijkl})\prod_{l=1}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{rtij2l}=y_{rtij2l}\mid\pi_{rtij2l})$$

Furthermore, it holds that:

$$\begin{aligned}
&P\Bigg(\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}(Y_{rtij2l}=y_{rtij2l})\bigcap_{l=1}^{f}\bigcap_{j=1}^{d}\bigcap_{i=1}^{c_j}\bigcap_{k\leq e,k\neq 2}(Y_{tijkl}=y_{tijkl})\;\Bigg|\;\mathcal{I}_{t111},\ldots,\mathcal{I}_{tc_dd1},\mathcal{S}_{t111}\ldots,\\
&\quad\mathcal{S}_{tc_dd1},\zeta_{t1111},\ldots,\zeta_{c_dd1f},\mathcal{I}^{M}_{t113},\ldots,\mathcal{I}^{M}_{tc_dde},\mathcal{S}^{M}_{t113},\ldots,\mathcal{S}^{M}_{tc_dde},\\
&\quad\zeta^{M}_{t131},\ldots,\zeta^{M}_{tdef},\mathcal{I}^{CM}_{t112},\ldots,\mathcal{I}^{CM}_{tc_dd2},\mathcal{S}^{CM}_{t112},\ldots,\mathcal{S}^{CM}_{tc_dd2},\zeta^{CM}_{t121},\ldots,\\
&\quad\zeta^{CM}_{td2f},\mathcal{I}^{UM}_{rt112},\ldots,\mathcal{I}^{UM}_{rtc_dd2},\mathcal{S}^{UM}_{rt112},\ldots,\mathcal{S}^{UM}_{rtc_dd2},\zeta^{UM}_{rt121},\ldots,\zeta^{UM}_{rtd2f}\Bigg)\\
&=\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{tij11}=y_{tij11}\mid\mathcal{I}_{tij1},\zeta_{tij11})\prod_{l=2}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{tij1l}=y_{tij1l}\mid\mathcal{I}_{tij1},\mathcal{S}_{tij1},\zeta_{tij1l})\\
&\quad\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{rtij21}=y_{rtij21}\mid\mathcal{I}_{tij1},\zeta_{tij1l},\mathcal{I}^{CM}_{tij2},\zeta^{CM}_{tj2l},\mathcal{I}^{UM}_{rtij2},\zeta^{UM}_{rtj2l})\\
&\quad\prod_{l=2}^{f}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{rtij2l}=y_{rtij2l}\mid\mathcal{I}_{tij1},\mathcal{S}_{tij1},\zeta_{tij1l},\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{tij2},\zeta^{CM}_{tj2l},\mathcal{I}^{UM}_{rtij2},\mathcal{S}^{UM}_{rtij2},\zeta^{UM}_{rtj2l})\\
&\quad\prod_{k=3}^{e}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{tijk1}=y_{tijk1}\mid\mathcal{I}_{tij1},\zeta_{tij1l},\mathcal{I}^{M}_{tijk},\zeta^{M}_{tjkl})\\
&\quad\prod_{l=2}^{f}\prod_{k=3}^{e}\prod_{j=1}^{d}\prod_{i=1}^{c_j}P(Y_{tijkl}=y_{tijkl}\mid\mathcal{I}_{tij1},\mathcal{S}_{tij1},\zeta_{tij1l},\mathcal{I}^{M}_{tijk},\mathcal{S}^{M}_{tijk},\zeta^{M}_{tjkl})
\end{aligned} \tag{5.7.2}$$

Remarks. According to Equation (4.7.2), all observed variables Y_{rtij2l} and Y_{tijkl} are independent given the latent response variables π_{rtij2l} and π_{tijkl}. Note that this assumption and its implications do not differ from the LST-Com GRM. Equation (5.7.1) implies that all associations between the observed variables are determined by the latent variables π_{tijkl} and π_{rtij2l} and their associations. According to Equation (5.7.2), the same holds with respect to the variables $\mathcal{I}_{tij1}$, $\mathcal{S}_{tij1}$, ζ_{tij1l}, $\mathcal{I}^{UM}_{rtij2}$, $\mathcal{S}^{UM}_{rtij2}$, ζ^{UM}_{rtj2l}, $\mathcal{I}^{CM}_{tij2}$, $\mathcal{S}^{CM}_{tij2}$, ζ^{CM}_{tj2l}, $\mathcal{I}^{M}_{tijk}$, $\mathcal{S}^{M}_{tijk}$, and ζ^{M}_{tjkl}.
As Equation (5.7.1) is identical to (4.7.2) in the LST-Com GRM it also follows from Equations (2.9.1) and (2.9.2) - (2.9.4). This is the case as the random variables π_{tij1l}, π_{tijkl}, $k > 2$, and π_{rtij2l} are (p_T, p_{TS_l})-, (p_T, p_{TS_l}, $p_{R_kS_l}$)-, and (p_T, p_{TS_l}, p_R, $p_{R_2S_l}$)-measurable functions, respectively. Similar arguments lead to Equation (4.7.3). A prove was given by Eid (1995, pp. 97-98) for a comparable model

and is applicable to the present case.

5.7.2 Conditional regressive independence of the latent state variables

Identical to the LST-Com GRM, independence assumption (4.7.1) of Definition 4.2 allows to interpret the latent unique method trait variables ξ^{UM}_{rtij2l} as the difference between the conditional expectation of the latent response variables π_{rtij2l} given the target and the rater and its conditional expectation given the target only (Koch, 2013).

Theorem 5.4. ***(LGC-Com GRM with conditional regressive independent latent state variables)***

$\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ be an LGC-Com GRM with (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables as defined by Definition 5.1. $\mathcal{M}$ is called LGC-Com GRM with conditionally regressive independent S_{tij2l} variables, if assumption (4.7.1) *of Definition 4.2 holds. Then, it follows that*

$$\mathbb{E}[S_{tij2l} \mid p_T, p_R] = \mathbb{E}[S_{tij2l} \mid p_T], \tag{5.7.3}$$

and the variables ξ^{UM}_{rtij2l} can be redefined as follows:

$$\xi^{UM}_{rtij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_R] - \mathbb{E}[\pi_{rtij2l} \mid p_T]. \tag{5.7.4}$$

Remarks. As in the LST-Com GRM, Theorem 5.4 states that, given the assumption given in Equation (5.7.3), the latent unique method trait variables ξ^{UM}_{rtij2l} can be interpreted as the difference between the conditional expectation of the latent response variables π_{rtij2l} given the target and the rater and its conditional expectation given the target only (see Section 4.7.2 for further details). As the latent unique method intercept and slope variables are direct functions of the latent unique method trait variables, this property applies to $\mathcal{I}^{UM}_{rtij2}$ and $\mathcal{S}^{UM}_{rtij2}$ as well.

5.7.3 Zero correlations based on the model definition

The LGC-Com GRM with conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}. The following theorem introduces the covariances that are zero as a result of the conditional independence assumptions.

Theorem 5.5. ***(Testability)***

If $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{I}}^{\mathbf{M}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t$, $\boldsymbol{\mathcal{S}}^{\mathbf{M}}_t$, $\boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_t$, $\boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ is an LGC-Com GRM with (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables and conditional independence, then, for all $i, i' \in I_j$, $j, j' \in J$, $k \in K$, and $l, l' \in L$, it holds that

1. *The latent intercept and slope variables are uncorrelated with the latent method intercept*

and slope variables:

$$Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{tij2}) = 0 \quad (5.7.5)$$
$$Cov(\mathcal{I}_{tij1}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \quad (5.7.6)$$
$$Cov(\mathcal{I}_{tij1}, \mathcal{I}^{M}_{tijk}) = 0 \qquad k > 2 \quad (5.7.7)$$

$$Cov(\mathcal{S}_{tij1}, \mathcal{S}^{CM}_{tij2}) = 0 \quad (5.7.8)$$
$$Cov(\mathcal{S}_{tij1}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \quad (5.7.9)$$
$$Cov(\mathcal{S}_{tij1}, \mathcal{S}^{M}_{tijk}) = 0 \qquad k > 2 \quad (5.7.10)$$

$$Cov(\mathcal{I}_{tij1}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \quad (5.7.11)$$
$$Cov(\mathcal{S}_{tij1}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \quad (5.7.12)$$

2. *The latent state residual variables are uncorrelated with the latent state residual method variables:*

$$Cov(\zeta_{tij1l}, \zeta^{CM}_{tj2l}) = 0 \quad (5.7.13)$$
$$Cov(\zeta_{tij1l}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.14)$$
$$Cov(\zeta_{tij1l}, \zeta^{M}_{tjkl}) = 0 \qquad k > 2 \quad (5.7.15)$$

3. *The latent intercept and slope variables are uncorrelated with all latent state residual (method) variables:*

$$Cov(\mathcal{I}_{tij1}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.16)$$
$$Cov(\mathcal{I}_{tij1}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.17)$$
$$Cov(\mathcal{I}_{tij1}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.18)$$
$$Cov(\mathcal{I}_{tij1}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.19)$$

$$Cov(\mathcal{S}_{tij1}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.20)$$
$$Cov(\mathcal{S}_{tij1}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.21)$$
$$Cov(\mathcal{S}_{tij1}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.22)$$
$$Cov(\mathcal{S}_{tij1}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.23)$$

4. *The latent method intercept and slope variables are uncorrelated with all latent state residual (method) variables:*

$$Cov(\mathcal{I}^{CM}_{tij2}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.24)$$
$$Cov(\mathcal{I}^{CM}_{tij2}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.25)$$
$$Cov(\mathcal{I}^{CM}_{tij2}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.26)$$
$$Cov(\mathcal{I}^{CM}_{tij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.27)$$

$$Cov(\mathcal{S}^{CM}_{tij2}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.28)$$
$$Cov(\mathcal{S}^{CM}_{tij2}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.29)$$
$$Cov(\mathcal{S}^{CM}_{tij2}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.30)$$
$$Cov(\mathcal{S}^{CM}_{tij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.31)$$

$$Cov(\mathcal{I}^{UM}_{rtij2}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.32)$$
$$Cov(\mathcal{I}^{UM}_{rtij2}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.33)$$
$$Cov(\mathcal{I}^{UM}_{rtij2}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.34)$$
$$Cov(\mathcal{I}^{UM}_{rtij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.35)$$

$$Cov(\mathcal{S}^{UM}_{rtij2}, \zeta_{ti'j'kl'}) = 0 \quad (5.7.36)$$
$$Cov(\mathcal{S}^{UM}_{rtij2}, \zeta^{CM}_{tj'2l'}) = 0 \quad (5.7.37)$$
$$Cov(\mathcal{S}^{UM}_{rtij2}, \zeta^{UM}_{rtj'2l'}) = 0 \quad (5.7.38)$$
$$Cov(\mathcal{S}^{UM}_{rtij2}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.39)$$

$$Cov(\mathcal{I}^{M}_{tijk}, \zeta_{ti'j'kl'}) = 0 \qquad k > 2 \quad (5.7.40)$$
$$Cov(\mathcal{I}^{M}_{tijk}, \zeta^{CM}_{tj'2l'}) = 0 \qquad k > 2 \quad (5.7.41)$$
$$Cov(\mathcal{I}^{M}_{tijk}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad k > 2 \quad (5.7.42)$$
$$Cov(\mathcal{I}^{M}_{tijk}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.43)$$

$$Cov(\mathcal{S}^{M}_{tijk}, \zeta_{ti'j'kl'}) = 0 \qquad k > 2 \quad (5.7.44)$$
$$Cov(\mathcal{S}^{M}_{tijk}, \zeta^{CM}_{tj'2l'}) = 0 \qquad k > 2 \quad (5.7.45)$$
$$Cov(\mathcal{S}^{M}_{tijk}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad k > 2 \quad (5.7.46)$$
$$Cov(\mathcal{S}^{M}_{tijk}, \zeta^{M}_{tj'kl'}) = 0 \qquad k > 2 \quad (5.7.47)$$

5. *Uncorrelatedness of latent method intercept and slope variables:*

$$Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \quad (5.7.48)$$
$$Cov(\mathcal{I}^{M}_{tijk}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \qquad k > 2 \quad (5.7.49)$$

$$Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \quad (5.7.50)$$
$$Cov(\mathcal{I}^{M}_{tijk}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \qquad k > 2 \quad (5.7.51)$$

$$Cov(\mathcal{S}^{CM}_{tij2}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \quad (5.7.52)$$
$$Cov(\mathcal{S}^{M}_{tijk}, \mathcal{S}^{UM}_{rti'j'2}) = 0 \qquad k > 2 \quad (5.7.53)$$

$$Cov(\mathcal{S}^{CM}_{tij2}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \quad (5.7.54)$$
$$Cov(\mathcal{S}^{M}_{tijk}, \mathcal{I}^{UM}_{rti'j'2}) = 0 \qquad k > 2 \quad (5.7.55)$$

6. *Uncorrelatedness of latent state residual method variables:*

$$Cov(\zeta^{CM}_{tj2l}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad (5.7.56)$$
$$Cov(\zeta^{M}_{tjkl}, \zeta^{UM}_{rtj'2l'}) = 0 \qquad k > 2 \qquad (5.7.57)$$

Proofs.

5.7.5 By Equation (5.2.6) $\mathcal{I}^{CM}_{tij2}$ *is defined as* $\mathcal{I}_{tij2} - \mathbb{E}[\mathcal{I}_{tij2} \mid \mathcal{I}_{tij1}]$. *Hence,* $\mathcal{I}^{CM}_{tij2}$ *is defined as a residual with respect to* $\mathcal{I}_{tij1}$. *As residuals are uncorrelated with their regressors, it follows that, for the same indicator i and construct j,* $Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{tij2}) = 0$.

5.7.6 As shown in Proof 4.8.2 $\xi^{UM}_{rti'j'2l}$ *is defined as a residual with respect to any* p_T*-measurable function. As* $\mathcal{I}^{UM}_{rti'j'2}$ *is defined as* $\xi^{UM}_{rti'j'21}$ *by Equation* (5.2.12) *and* $\mathcal{I}_{tij1}$ *is defined as* $\mathcal{I}_{tij1} = \xi_{tij11}$*, and thereby a* p_T*-measurable function, it follows that* $Cov(\mathcal{I}_{tij1}, \mathcal{I}^{UM}_{rti'j'2}) = 0$.

5.7.7 The proof of Equation (5.7.7) *follows the same logic as Proof 5.7.5.*

5.7.8 By Equation (5.2.8) $\mathcal{S}^{CM}_{tij2}$ *is defined as* $\mathcal{S}_{tij2} - \mathbb{E}[\mathcal{S}_{tij2} \mid \mathcal{S}_{tij1}]$. *Hence,* $\mathcal{S}^{CM}_{tij2}$ *is defined as a residual with respect to* $\mathcal{S}_{tij1}$. *As residuals are uncorrelated with their regressors, it follows that, for the same indicator i and construct j,* $Cov(\mathcal{S}_{tij1}, \mathcal{S}^{CM}_{tij2}) = 0$.

5.7.9 As shown in Proof 4.8.2 $\xi^{UM}_{rti'j'2l}$ *is defined as a residual with respect to any* p_T*-measurable function. As* $\mathcal{S}^{UM}_{rti'j'2}$ *is a direct function of* $\xi^{UM}_{rti'j'21}$ *and* $\xi^{UM}_{rti'j'22}$ *by Equation* (5.2.13) *and* $\mathcal{S}_{tij1}$ *is a direct function of* ξ_{tij11} *and* ξ_{tij12} *by Equation* (5.2.5)*, it follows that* $Cov(\mathcal{S}_{tij1}, \mathcal{S}^{UM}_{rti'j'2}) = 0$.

5.7.10 The proof of Equation (5.7.10) *follows the same logic as Proof 5.7.8.*

5.7.11 The proof of Equation (5.7.11) *follows the same logic as Proofs 5.7.6 and 5.7.9.*

5.7.12 The proof of Equation (5.7.12) *follows the same logic as Proofs 5.7.6 and 5.7.9.*

The zero correlations given in Equations (5.7.13) - (5.7.15) *and* (5.7.56) - (5.7.57) *appear identically in the LST-Com GRM and were proven in Proofs 4.8.4 - 4.8.6 and 4.8.25 - 4.8.26.*

The zero correlations given in Equations (5.7.16) - (5.7.23) *follow from Proofs 4.8.7 - 4.8.10 in Section 4.8, as* $\mathcal{I}_{tij1}$ *and* $\mathcal{S}_{tij1}$ *are direct functions of* ξ_{tij1l}.

The zero correlations given in Equations (5.7.24) - (5.7.47) *follow from Proofs 4.8.11 - 4.8.22 in Section 4.8, as the variables* $\mathcal{I}^{UM}_{rtij2}$ *and* $\mathcal{S}^{UM}_{rtij2}$ *are direct functions of the variables* ξ^{UM}_{rtij2l}*, and the variables* $\mathcal{I}^{CM}_{tij2}$*,* $\mathcal{I}^{M}_{tijk}$*,* $\mathcal{S}^{CM}_{tij2}$*, and* $\mathcal{S}^{M}_{tijk}$ *are, like the latent trait method variables* ξ^{CM}_{tij2} *and* ξ^{M}_{tijk}*, direct functions of* ξ_{tijkl} *and* ξ_{tij1l}.

The zero correlations given in Equations (5.7.48) - (5.7.55) *follow from Proofs 4.8.23 - 4.8.24 as the variables* $\mathcal{I}^{UM}_{rtij2}$ *and* $\mathcal{S}^{UM}_{rtij2}$ *are direct functions of the variables* ξ^{UM}_{rtij2l}*, and the variables* $\mathcal{I}^{CM}_{tij2}$*,* $\mathcal{I}^{M}_{tijk}$*,* $\mathcal{S}^{CM}_{tij2}$*, and* $\mathcal{S}^{M}_{tijk}$ *are, like the latent trait method variables* ξ^{CM}_{tij2} *and* ξ^{M}_{tijk}*, direct functions of* ξ_{tijkl} *and* ξ_{tij1l}.

Remarks. Note that, as in the LST-Com GRM, the conditional independence assumptions given in Definition 5.2 do not imply that the latent state residuals ζ_{tij1l} or the latent state residual method

factors ζ^{CM}_{tj2l}, ζ^{UM}_{rtj2l}, and ζ^{M}_{tjkl} are uncorrelated over time. As in the LST-Com GRM, it is necessary to set some or all of these correlations to zero for identification reasons. Nevertheless, the model in its current form does for example allow to model autoregressive processes on the level of the state residual (method) variables. For details and a discussion see Section 4.8.
The following additional conditional independence assumption would lead to uncorrelated latent state residual and latent state residual method variables.

Definition 5.3. (LGC-Com GRM with strong conditional independence)

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\mathcal{I}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\mathcal{S}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\mathcal{I}^{\mathbf{UM}}_{rt}$, $\mathcal{I}^{\mathbf{CM}}_{t}$, $\mathcal{I}^{\mathbf{M}}_{t}$, $\mathcal{S}^{\mathbf{UM}}_{rt}$, $\mathcal{S}^{\mathbf{CM}}_{t}$, $\mathcal{S}^{\mathbf{M}}_{t}$, $\boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_{t}$, $\boldsymbol{\zeta}^{\mathbf{M}}_{t}\rangle$ *be an LGC-Com GRM with* (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})*-congeneric variables.* $\mathcal{M}$ *is called LGC-Com GRM with strong conditional independence if and only if the assumptions given in Definition 5.2 and Definition 4.3 hold.*

The definition of strong conditional independence in the LGC-Com GRM is identical to the assumptions made for strong conditional independence in the LST-Com GRM (see Definition 4.3) and thereby has the same meaning and implications, stated in the following theorem.

Theorem 5.6. ***(Testability of LGC-Com GRM with strong conditional independence)***

If $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\mathcal{I}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\mathcal{S}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\mathcal{I}^{\mathbf{UM}}_{rt}$, $\mathcal{I}^{\mathbf{CM}}_{t}$, $\mathcal{I}^{\mathbf{M}}_{t}$, $\mathcal{S}^{\mathbf{UM}}_{rt}$, $\mathcal{S}^{\mathbf{CM}}_{t}$, $\mathcal{S}^{\mathbf{M}}_{t}$, $\boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}$, $\boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}$, $\boldsymbol{\zeta}^{\mathbf{UM}}_{rt}$, $\boldsymbol{\zeta}^{\mathbf{CM}}_{t}$, $\boldsymbol{\zeta}^{\mathbf{M}}_{t}\rangle$ *is an LGC-Com GRM with* (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})*-congeneric variables and strong conditional independence, then Equations* (5.7.5) - (5.7.57) *of Theorem 5.5 hold and Equations* (4.8.29) - (4.8.35) *of Theorem 4.6 hold.*

Proofs and a discussion are provided in Section 4.8.

5.7.4 Covariance structure

The LGC-Com GRM with strong conditional independence implies a specific covariance structure of the latent variables π_{tijkl}, $k \neq 2$, and π_{rtij2l}, including the zero-correlations specified in Theorem 5.5 and 5.6. Whether this covariance structure holds in empirical applications can be tested based on the covariance structure of the variables Y^*_{tijkl} and Y^*_{rtij2l}, as defined in Section 2.8, with SEMs for ordinal observed variables (also see Section 4.8).
The total covariance matrix Σ_T of the variables Y^*_{tijkl} and Y^*_{rtij2l} in an LGC-Com GRM with strong conditional independence can be partitioned, just as in the LST-Com GRM, into a within and a between covariance matrix and can be represented as

$$\Sigma_T = \Lambda_{\xi B}\Phi_{\xi B}\Lambda'_{\xi B} + \Lambda_{\zeta B}\Phi_{\zeta B}\Lambda'_{\zeta B} + \Theta_B + \Lambda_{\xi W}\Phi_{\xi W}\Lambda'_{\xi W} + \Lambda_{\zeta W}\Phi_{\zeta W}\Lambda'_{\zeta W} + \Theta_W \qquad (5.7.58)$$

where $\Lambda_{\xi B}$ and $\Lambda_{\xi W}$ refer to the factor loading matrices of the trait-specific variables (intercept and slope factors) on the between- and within-level, respectively, $\Lambda_{\zeta B}$ and $\Lambda_{\zeta W}$ refer to the factor loading matrices of the occasion-specific variables (state residual and method state residual variables) on the between- and within-level, respectively, $\Phi_{\xi B}$ and $\Phi_{\xi W}$ refer to the variance-covariance matrices of the between and within trait-specific latent variables, respectively, $\Phi_{\zeta B}$ and $\Phi_{\zeta W}$ refer to the variance-covariance matrices of the between and within occasion-specific latent variables, respectively, and Θ_B and Θ_W are the between- and within-level residual variance-covariance matrices.

The structure of the covariance matrices $\Phi_{\zeta W}$ and $\Phi_{\zeta B}$ and their factor loading matrices $\Lambda_{\zeta W}$ and $\Lambda_{\zeta B}$ equal those of the LST-Com GRM, and thereby the LST-Com model for continuous indicator variables given in Koch et al. (2017, supplementary material, pp. 13-20). The residual variance-covariance matrices Θ_B and Θ_W equal those of the LST-Com model for continuous indicator variables with the exception that all non-zero elements $Var(E_{rtij2l})$ and $Var(E_{tijkl})$ in the matrices Θ_B and Θ_W have to be replaced by 1.
The structure of the covariance matrices $\Phi_{\xi W}$ and $\Phi_{\xi B}$ and their factor loading matrices $\Lambda_{\xi W}$ and $\Lambda_{\xi B}$ differ from those of the LST-Com GRM. Note that the structure of $\Phi_{\xi W}$ and $\Phi_{\xi B}$ also differs from those of the continuous indicator LGC-Com model as defined by Koch (2013), as the LGC model was defined in a different way. In the following, the covariance matrices $\Phi_{\xi W}$ and $\Phi_{\xi B}$ and their factor loading matrices $\Lambda_{\xi W}$ and $\Lambda_{\xi B}$ for an LGC-Com GRM with strong conditional independence as defined in Definition 5.2 and 5.3 are described for a model with three indicators ($i_j = 3\ \forall j$), two traits ($j = 2$), two methods ($k = 2$, one structurally different reference method and one set of interchangeable methods), and three occasions of measurement ($l = 3$).

Let the vector of the latent response variables Y^* be given by

$$\begin{aligned}\mathbf{Y}^* = \big(&Y^*_{t1111}, Y^*_{t2111}, Y^*_{t3111}, Y^*_{t1121}, Y^*_{t2121}, Y^*_{t3121}, Y^*_{t1112}, Y^*_{t2112}, Y^*_{t3112}, Y^*_{t1122}, Y^*_{t2122}, Y^*_{t3122},\\ &Y^*_{t1113}, Y^*_{t2113}, Y^*_{t3113}, Y^*_{t1123}, Y^*_{t2123}, Y^*_{t3123}, Y^*_{t1211}, Y^*_{t2211}, Y^*_{t3211}, Y^*_{t1221}, Y^*_{t2221}, Y^*_{t3221},\\ &Y^*_{t1212}, Y^*_{t2212}, Y^*_{t31212}, Y^*_{t1222}, Y^*_{t2222}, Y^*_{t3222}, Y^*_{t1213}, Y^*_{t2213}, Y^*_{t3213}, Y^*_{t1223}, Y^*_{t2223}, Y^*_{t3223}\big)'\end{aligned}$$

with

$$\Sigma_T = \mathbb{E}\big[\big(\mathbf{Y}^* - \mathbb{E}[\mathbf{Y}^*]\big)\big(\mathbf{Y}^* - \mathbb{E}[\mathbf{Y}^*]\big)'\big].$$

Let $\Phi_{\xi B}$ be given by

$$\Phi_{\xi B} = \mathbb{E}\big[\big(\mathbf{V}_{\xi\mathbf{B}} - \mathbb{E}[\mathbf{V}_{\xi\mathbf{B}}]\big)\big(\mathbf{V}_{\xi\mathbf{B}} - \mathbb{E}[\mathbf{V}_{\xi\mathbf{B}}]\big)'\big]$$

with

$$\begin{aligned}\mathbf{V}_{\xi\mathbf{B}} = \big(&\mathcal{I}_{t111}, \mathcal{I}_{t211}, \mathcal{I}_{t311}, \mathcal{S}_{t111}, \mathcal{S}_{t211}, \mathcal{S}_{t311},\\ &\mathcal{I}^{CM}_{t112}, \mathcal{I}^{CM}_{t212}, \mathcal{I}^{CM}_{t312}, \mathcal{S}^{CM}_{t112}, \mathcal{S}^{CM}_{t212}, \mathcal{S}^{CM}_{t312},\\ &\mathcal{I}_{t121}, \mathcal{I}_{t221}, \mathcal{I}_{t321}, \mathcal{S}_{t121}, \mathcal{S}_{t221}, \mathcal{S}_{t321},\\ &\mathcal{I}^{CM}_{t122}, \mathcal{I}^{CM}_{t222}, \mathcal{I}^{CM}_{t322}, \mathcal{S}^{CM}_{t122}, \mathcal{S}^{CM}_{t222}, \mathcal{S}^{CM}_{t322}\big)'\end{aligned}$$

and $\Phi_{\xi W}$ by

$$\Phi_{\xi W} = \mathbb{E}\big[\big(\mathbf{V}_{\xi\mathbf{W}} - \mathbb{E}[\mathbf{V}_{\xi\mathbf{W}}]\big)\big(\mathbf{V}_{\xi\mathbf{W}} - \mathbb{E}[\mathbf{V}_{\xi\mathbf{W}}]\big)'\big]$$

with

$$\begin{aligned}\mathbf{V}_{\xi\mathbf{W}} = \big(&\mathcal{I}^{UM}_{t112}, \mathcal{I}^{UM}_{t212}, \mathcal{I}^{UM}_{t312}, \mathcal{S}^{UM}_{t112}, \mathcal{S}^{UM}_{t212}, \mathcal{S}^{UM}_{t312},\\ &\mathcal{I}^{UM}_{t122}, \mathcal{I}^{UM}_{t222}, \mathcal{I}^{UM}_{t322}, \mathcal{S}^{UM}_{t122}, \mathcal{S}^{UM}_{t222}, \mathcal{S}^{UM}_{t322}\big)'\end{aligned}$$

Then the matrix of the between-level latent intercept and slope factor loadings $\Lambda_{\xi B}$ is given by

$$\Lambda_{\xi B} = \sum_{p=1}^{6} I^p_{\Lambda_\xi} \otimes \Lambda_{\xi B_p}$$

with

$$I^1_{\Lambda_\xi} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad I^2_{\Lambda_\xi} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad I^3_{\Lambda_\xi} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$I^4_{\Lambda_\xi} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad I^5_{\Lambda_\xi} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad I^6_{\Lambda_\xi} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\Lambda_{\xi B_p} = \begin{pmatrix} 1 & 0 & 0 & (l-1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & (l-1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & (l-1) & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{\mathcal{I}1j2} & 0 & 0 & (l-1)\lambda_{\mathcal{S}1j2} & 0 & 0 & 1 & 0 & 0 & (l-1) & 0 & 0 \\ 0 & \lambda_{\mathcal{I}2j2} & 0 & 0 & (l-1)\lambda_{\mathcal{I}2j2} & 0 & 0 & 1 & 0 & 0 & (l-1) & 0 \\ 0 & 0 & \lambda_{\mathcal{I}3j2} & 0 & 0 & (l-1)\lambda_{\mathcal{I}3j2} & 0 & 0 & 1 & 0 & 0 & (l-1) \end{pmatrix}$$

where $\otimes$ denotes the Kronecker product and $l = 1$ for $p \in \{1,4\}$, $l = 2$ for $p \in \{2,5\}$, $l = 3$ for $p \in \{3,6\}$, and $j = 1$ for $p \in \{1,2,3\}$, and $j = 2$ for $p \in \{4,5,6\}$.

The matrix of the within-level latent intercept and slope factor loadings $\Lambda_{\xi W}$ is given by

$$\Lambda_{\xi W} = \sum_{p=1}^{6} I^p_{\Lambda_\xi} \otimes \Lambda_{\xi W_p}$$

with

$$\Lambda_{\xi W_p} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & (l-1) & 0 & 0 \\ 0 & 1 & 0 & 0 & (l-1) & 0 \\ 0 & 0 & 1 & 0 & 0 & (l-1) \end{pmatrix}$$

and $l = 1$ for $p \in \{1,4\}$, $l = 2$ for $p \in \{2,5\}$, $l = 3$ for $p \in \{3,6\}$, and $I^p_{\Lambda_\xi}$ as defined above.

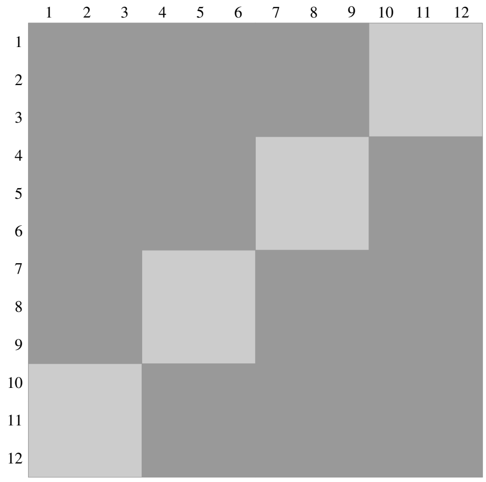

Figure 5.3: Within-level variance-covariance matrix $\Phi_{\xi W}$ of the LGC-Com GRM, where 1=$\mathcal{I}^{UM}_{t112}$, 2=$\mathcal{I}^{UM}_{t212}$, 3=$\mathcal{I}^{UM}_{t312}$, 4=$\mathcal{S}^{UM}_{t112}$, 5=$\mathcal{S}^{UM}_{t212}$, 6=$\mathcal{S}^{UM}_{t312}$, 7=$\mathcal{I}^{UM}_{t122}$, 8=$\mathcal{I}^{UM}_{t222}$, 9=$\mathcal{I}^{UM}_{t322}$, 10=$\mathcal{S}^{UM}_{t122}$, 11=$\mathcal{S}^{UM}_{t222}$, 12=$\mathcal{S}^{UM}_{t322}$. Cells colored in gray indicate permissible and interpretable correlations. Cells in light gray indicate correlations that could be fixed to zero for parsimony.

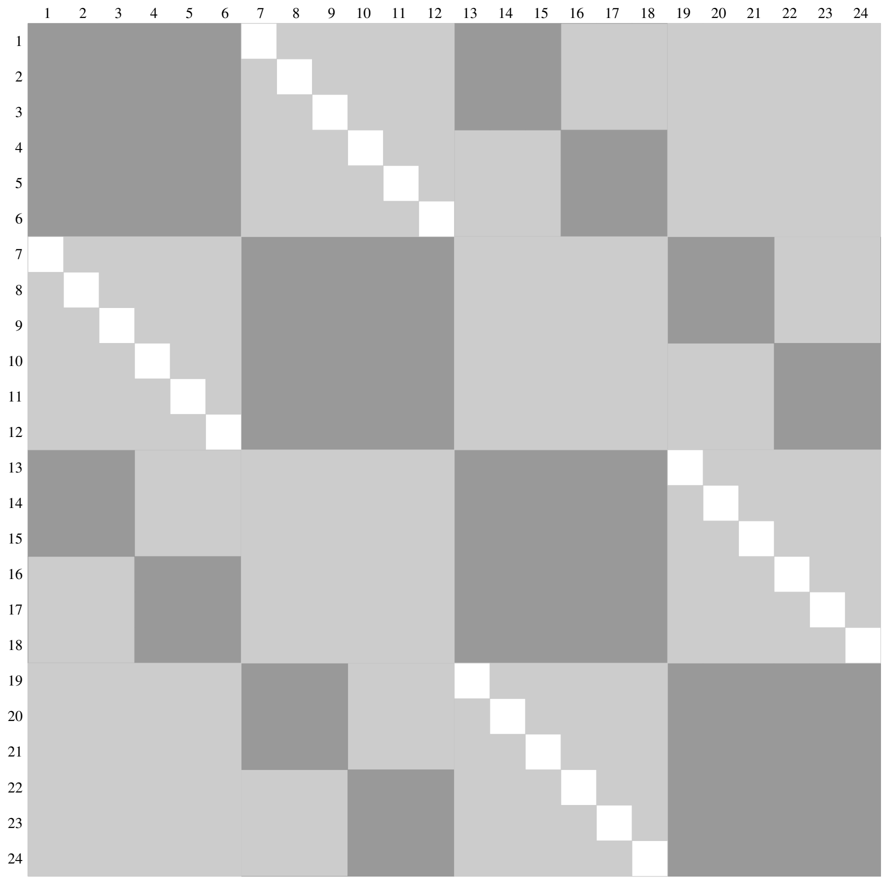

Figure 5.4: Between variance-covariance matrix $\Phi_{\xi B}$ of the LGC-Com GRM, where 1=$\mathcal{I}_{t111}$, 2=$\mathcal{I}_{t211}$, 3=$\mathcal{I}_{t311}$, 4=$\mathcal{S}_{t111}$, 5=$\mathcal{S}_{t211}$, 6=$\mathcal{S}_{t311}$, 7=$\mathcal{I}^{CM}_{t112}$, 8=$\mathcal{I}^{CM}_{t212}$, 9=$\mathcal{I}^{CM}_{t312}$, 10=$\mathcal{S}^{CM}_{t112}$, 11=$\mathcal{S}^{CM}_{t212}$, 12=$\mathcal{S}^{CM}_{t312}$, 13=$\mathcal{I}_{t121}$, 14=$\mathcal{I}_{t221}$, 15=$\mathcal{I}_{t321}$, 16=$\mathcal{S}_{t121}$, 17=$\mathcal{S}_{t221}$, 18=$\mathcal{S}_{t321}$, 19=$\mathcal{I}^{CM}_{t122}$, 20=$\mathcal{I}^{CM}_{t222}$, 21=$\mathcal{I}^{CM}_{t322}$, 22=$\mathcal{S}^{CM}_{t122}$, 23=$\mathcal{S}^{CM}_{t222}$, 24=$\mathcal{S}^{CM}_{t322}$. Cells colored in white indicate zero correlations, cells colored in gray indicate permissible and interpretable correlations. Cells in light gray indicate correlations that could be fixed to zero for parsimony.

Figure 5.3 illustrates the latent variance-covariance Φ_{ξ_W} and Figure 5.4 illustrates the the latent variance-covariance matrix Φ_{ξ_B} of the LGC-Com GRM with strong conditional independence. Note that all covariances between latent unique method intercept and slope variables are permissible. For a detailed description of the occasion-specific variance-covariance matrices Φ_{ζ_W} and Φ_{ζ_B} and their factor loading matrices Λ_{ζ_W} and Λ_{ζ_B} see Koch et al. (2017).

5.8 Variance decompositions

Based on the definition of the LGC-Com GRM, the latent response variables π_{tijkl} and π_{rtij2l} can be additively decomposed into different variance components. From Definition 5.1 and Theorem 5.1 it follows that the general measurement equations for the latent response variables in an LGC-Com GRM of (ζ^{CM}_{tij2l}, ζ^{UM}_{tij2l}, ζ^{M}_{tijkl})-congeneric variables are given by:

$$\pi_{tij1l} = \mathcal{I}_{tij1} + (l-1)\mathcal{S}_{tij1} + \zeta_{tij1l} \tag{5.8.1}$$

$$\begin{aligned}\pi_{tijkl} = {} & \alpha_{\mathcal{I}ijk} + \alpha_{\mathcal{S}ijk} + \lambda_{\mathcal{I}ijk}\mathcal{I}_{tij1} + (l-1)\lambda_{\mathcal{S}ijk}\mathcal{S}_{tij1} \\ & + \mathcal{I}^{M}_{tijk} + (l-1)\mathcal{S}^{M}_{tijk} + \lambda_{\zeta ijkl}\zeta_{tij1l} + \lambda^{M}_{\zeta ijkl}\zeta^{M}_{tjkl} \qquad k>2\end{aligned} \tag{5.8.2}$$

$$\begin{aligned}\pi_{rtij2l} = {} & \alpha_{\mathcal{I}ij2} + \alpha_{\mathcal{S}ij2} + \lambda_{\mathcal{I}ij2}\mathcal{I}_{tij1} + (l-1)\lambda_{\mathcal{S}ij2}\mathcal{S}_{tij1} \\ & + \mathcal{I}^{CM}_{tij2} + (l-1)\mathcal{S}^{CM}_{tij2} + \mathcal{I}^{UM}_{rtij2} + (l-1)\mathcal{S}^{UM}_{rtij2} \\ & + \lambda_{\zeta ij2l}\zeta_{tij1l} + \lambda^{CM}_{\zeta ij2l}\zeta^{CM}_{tj2l} + \lambda^{UM}_{\zeta ij2l}\zeta^{UM}_{rtj2l}\end{aligned} \tag{5.8.3}$$

As the latent method variables are defined as latent residual variables, they are uncorrelated with their respective regressors. That is, due to the zero-covariances given in Equations (5.7.5) - (5.7.57), the different variance components can be separated. The variances of the latent response variables can therefore be additively decomposed as:

$$Var(\pi_{tij1l}) = Var(\mathcal{I}_{tij1}) + (l-1)^2 Var(\mathcal{S}_{tij1}) + (l-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) + Var(\zeta_{tij1l}) \tag{5.8.4}$$

$$\begin{aligned}Var(\pi_{tijkl}) = {} & (\lambda_{\mathcal{I}ijk})^2 Var(\mathcal{I}_{tij1}) + (l-1)^2(\lambda_{\mathcal{S}ijk})^2 Var(\mathcal{S}_{tij1}) \\ & + (l-1)\lambda_{\mathcal{I}ijk}\lambda_{\mathcal{S}ijk}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ & + Var(\mathcal{I}^{M}_{tijk}) + (l-1)^2 Var(\mathcal{S}^{M}_{tijk}) + (l-1)Cov(\mathcal{I}^{M}_{tijk}, \mathcal{S}^{M}_{tijk}) \\ & + (l-1)\lambda_{\mathcal{I}ijk}Cov(\mathcal{I}_{tij1}, \mathcal{S}^{M}_{tijk}) + (l-1)\lambda_{\mathcal{S}ijk}Cov(\mathcal{S}_{tij1}, \mathcal{I}^{M}_{tijk}) \\ & + (\lambda_{\zeta ijkl})^2 Var(\zeta_{tij1l}) + (\lambda^{M}_{\zeta ijkl})^2 Var(\zeta^{M}_{tjkl}) \qquad k>2\end{aligned} \tag{5.8.5}$$

$$\begin{aligned}Var(\pi_{rtij2l}) = {} & (\lambda_{\mathcal{I}ij2})^2 Var(\mathcal{I}_{tij1}) + (l-1)^2(\lambda_{\mathcal{S}ij2})^2 Var(\mathcal{S}_{tij1}) \\ & + (l-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ & + Var(\mathcal{I}^{CM}_{tij2}) + (l-1)^2 Var(\mathcal{S}^{CM}_{tij2}) + (l-1)Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{tij2}) \\ & + (l-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{tij2}) + (l-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) \\ & + Var(\mathcal{I}^{UM}_{rtij2}) + (l-1)^2 Var(\mathcal{S}^{UM}_{rtij2}) + (l-1)Cov(\mathcal{I}^{UM}_{rtij2}, (\mathcal{S}^{UM}_{rtij2}) \\ & + (\lambda_{\zeta ij2l})^2 Var(\zeta_{tij1l}) + (\lambda^{CM}_{\zeta ij2l})^2 Var(\zeta^{CM}_{tj2l}) + (\lambda^{UM}_{\zeta ij2l})^2 Var(\zeta^{UM}_{rtj2l})\end{aligned} \tag{5.8.6}$$

Based on this variance decomposition, different variance components can be defined. It should be noted that latent growth curve models implicitly assume that the variance of the latent response vari-

ables increase in a non-linear fashion as the number of measurement occasions increases above three. Although this is not necessarily the case with only three measurement occasions, it is also true if intercept and slope factors are positively correlated. Therefore, it was recommended not to compare different variance components over time points (Koch, 2013).
In general, the same variance components could be computed as in the LST-Com GRM. Definitions of these variance coefficients are given in Table 4.1, where the variance of the latent trait variables have to be replaced by their respective intercept and slope components. In addition, the variance components given in Table 5.1 could be computed. The growth consistency coefficients represent the amount of growth in the non-reference methods that can be explained by the growth in the reference-method indicators.

Table 5.1: Definition of the Growth Consistency coefficients in the LGC-Com GRM.

Growth Consistency Coefficients		
Method	Level	Definition
Interchangeable	Target	$GCon(\pi_{tij2l}) = \frac{(\lambda_{\mathcal{S}ij2})^2 Var(\mathcal{S}_{tij1})}{(\lambda_{\mathcal{S}ij2})^2 Var(\mathcal{S}_{tij1}) + Var(\mathcal{S}^{CM}_{tij2})}$
Interchangeable	Rater	$GCon(\pi_{rtij2l}) = \frac{(\lambda_{\mathcal{S}ij2})^2 Var(\mathcal{S}_{tij1})}{(\lambda_{\mathcal{S}ij2})^2 Var(\mathcal{S}_{tij1}) + Var(\mathcal{S}^{CM}_{tij2}) + Var(\mathcal{S}^{UM}_{rtij2})}$
Structurally different	Target	$GCon(\pi_{tijkl}) = \frac{(\lambda_{\mathcal{S}ijk})^2 Var(\mathcal{S}_{tij1})}{(\lambda_{\mathcal{S}ijk})^2 Var(\mathcal{S}_{tij1}) + Var(\mathcal{S}^{M}_{tij2})}$

5.9 Mean structure

The following theorem clarifies the mean structure of the latent variables in the LGC-Com GRM. The mean structure is needed to derive the identification conditions in Section 5.10.

Theorem 5.7. ***(Mean Structure)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P}), \boldsymbol{\pi}_{rt}, \boldsymbol{\pi}_t, \boldsymbol{\kappa}, \boldsymbol{\alpha}_{\mathcal{I}}, \boldsymbol{\lambda}_{\mathcal{I}}, \boldsymbol{\mathcal{I}}_t, \boldsymbol{\alpha}_{\mathcal{S}}, \boldsymbol{\lambda}_{\mathcal{S}}, \boldsymbol{\mathcal{S}}_t, \boldsymbol{\lambda}_{\zeta}, \boldsymbol{\zeta}_t, \boldsymbol{\mathcal{I}}^{\mathbf{UM}}_{rt}, \boldsymbol{\mathcal{I}}^{\mathbf{CM}}_t, \boldsymbol{\mathcal{I}}^{\mathbf{M}}_t, \boldsymbol{\mathcal{S}}^{\mathbf{UM}}_{rt}, \boldsymbol{\mathcal{S}}^{\mathbf{CM}}_t, \boldsymbol{\mathcal{S}}^{\mathbf{M}}_t, \boldsymbol{\lambda}^{\mathbf{UM}}_{\zeta}, \boldsymbol{\lambda}^{\mathbf{CM}}_{\zeta}, \boldsymbol{\lambda}^{\mathbf{M}}_{\zeta}, \boldsymbol{\zeta}^{\mathbf{UM}}_{rt}, \boldsymbol{\zeta}^{\mathbf{CM}}_t, \boldsymbol{\zeta}^{\mathbf{M}}_t\rangle$ *be an LGC-Com GRM with* $(\zeta^{CM}_{tij2l}, \zeta^{UM}_{tij2l}, \zeta^{M}_{tijkl})$*-congeneric variables and conditional independence. Then, for all* $j \in J$, $i \in I_j$, $k \in K$, *and* $l \in L$ *it holds that*

$$\mathbb{E}(\mathcal{I}^{UM}_{rtij2}) = 0 \tag{5.9.1}$$

$$\mathbb{E}(\mathcal{S}^{UM}_{rtij2}) = 0 \tag{5.9.2}$$

$$\mathbb{E}(\mathcal{I}^{CM}_{tij2}) = 0 \tag{5.9.3}$$

$$\mathbb{E}(\mathcal{S}^{CM}_{tij2}) = 0 \tag{5.9.4}$$

$$\mathbb{E}(\mathcal{I}^{M}_{tijk}) = 0 \qquad k > 2 \tag{5.9.5}$$

$$\mathbb{E}(\mathcal{S}^{M}_{tijk}) = 0 \qquad k > 2 \tag{5.9.6}$$

(5.9.7)

$$\mathbb{E}(\zeta^{UM}_{rtj2l}) = 0 \quad (5.9.8)$$

$$\mathbb{E}(\zeta^{CM}_{tj2l}) = 0 \quad (5.9.9)$$

$$\mathbb{E}(\zeta^{M}_{tjkl}) = 0 \qquad k > 2 \quad (5.9.10)$$

$$\mathbb{E}(\zeta_{tij1l}) = 0 \quad (5.9.11)$$

$$\mathbb{E}(\pi_{tsijkl}) = \mathbb{E}(\pi_{tijkl}) - \kappa_{sijkl} \qquad k \neq 2 \quad (5.9.12)$$

$$\mathbb{E}(\pi_{rtsij2l}) = \mathbb{E}(\pi_{rtij2l}) - \kappa_{sij2l} \quad (5.9.13)$$

$$\mathbb{E}(\pi_{tij11}) = \mathbb{E}(\mathcal{I}_{tij1}) \quad (5.9.14)$$

$$\mathbb{E}(\pi_{tij1l}) = \mathbb{E}(\mathcal{I}_{tij1}) + \mathbb{S}(\mathcal{I}_{tij1}) \quad (5.9.15)$$

$$\mathbb{E}(\pi_{(r)tijk1}) = \alpha_{\mathcal{I}ijk} + \lambda_{\mathcal{I}ijk}\, \mathbb{E}(\mathcal{I}_{tij1}) \qquad k > 1 \quad (5.9.16)$$

$$\mathbb{E}(\pi_{(r)tijkl}) = \alpha_{\mathcal{I}ijk} + \alpha_{\mathcal{S}ijk} + \lambda_{\mathcal{I}ijk}\, \mathbb{E}(\mathcal{I}_{tij1}) + \lambda_{\mathcal{S}ijk}\, \mathbb{E}(\mathcal{S}_{tij1}) \qquad k > 1 \quad (5.9.17)$$

and in LGC-Com GRMs defined with common latent intercept and slope factors $\mathcal{I}_{tj1}$, $\mathcal{S}_{tj1}$:

$$\mathbb{E}(\mathcal{I}_{tij1}) = \delta_{\mathcal{I}ij1} + \lambda_{\mathcal{I}ij1}\, \mathbb{E}(\mathcal{I}_{tj1}) \quad (5.9.18)$$

$$\mathbb{E}(\mathcal{S}_{tij1}) = \delta_{\mathcal{S}ij1} + \lambda_{\mathcal{S}ij1}\, \mathbb{E}(\mathcal{S}_{tj1}) \quad (5.9.19)$$

Proofs. *Mean Structure.*
Equations (5.9.1) - (5.9.11) *follow directly from the definition of the latent (trait and state residual) method variables as well as the latent state residual variables as residual variables in Definition 4.1 and the fact that residual variables have an expectation of zero (Steyer & Nagel, 2017, p. 323). Equations* (5.9.12) *and* (5.9.13) *follow directly from the definitions of the latent response variables π_{tsijkl} and $\pi_{rtsij2l}$ given in Definition 2.1. Equations* (5.9.14) - (5.9.17) *follow directly from Definition 5.1 as well as from Equations* (5.9.1) - (5.9.11). *Equations* (5.9.18) - (5.9.19) *follow from Equation* (5.4.2) *and their analogon for common latent slope variables. The proofs are straightforward and therefore left to the reader.*

Remarks. Equation (5.9.14) shows that the expected value of the latent intercept variable $\mathcal{I}_{tij1}$ equals the expectation of the latent response variable π_{tij11} . For models defined with common (non-indicator-specific) latent intercept factors $\mathcal{I}_{tj1}$, the expectation of the latent trait factors $\mathcal{I}_{tj1}$ equals the expectation of the latent response variables π_{tij11} if and only if $\delta_{\mathcal{I}ij1} = 0$ and $\lambda_{\mathcal{I}ij1} = 1$.

5.10 Identifiability

Theorem 5.8 defines identifiability conditions for the LGC-Com GRM with indicator-specific latent intercept and slope (method) variables $\mathcal{I}_{tij1}$, $\mathcal{S}_{tij1}$, $\mathcal{I}^{CM}_{tij2}$, $\mathcal{S}^{CM}_{tij2}$, $\mathcal{I}^{UM}_{rtij2}$, $\mathcal{S}^{UM}_{rtij2}$, $\mathcal{I}^{M}_{tijk}$, and $\mathcal{S}^{M}_{tijk}$. Identifiability conditions for the model with common (non-indicator-specific) intercept and slope (method) factors can easily be derived from the conditions given in Theorem 5.8. Furthermore, recall that all latent method factors have an expectation of zero by definition.

In Equation (5.7.58) the total covariance matrix of the variables Y^*_{tijkl} and Y^*_{rtij2l} in an LGC-Com GRM was represented as

$$\Sigma_T = \Lambda_{\xi B}\Phi_{\xi B}\Lambda'_{\xi B} + \Lambda_{\zeta B}\Phi_{\zeta B}\Lambda'_{\zeta B} + \Theta_B + \Lambda_{\xi W}\Phi_{\xi W}\Lambda'_{\xi W} + \Lambda_{\zeta W}\Phi_{\zeta W}\Lambda'_{\zeta W} + \Theta_W$$

where all non-zero elements in the residual variance-covariance matrices Θ_B and Θ_W are equal to 1. Theorem 5.8 then gives identification conditions for the LGC-Com GRM parameters. Note that all threshold parameters κ_{sijkl} are set invariant over measurement occasions by definition of the LGC-Com GRM given in Definition 5.1, that is, $\kappa_{sijkl} = \kappa_{sijkl'}\ \forall\, l, l' \in L$.

Theorem 5.8. ***(Identification of the LGC-Com GRM)***

Let $\mathcal{M} = \langle(\Omega, \mathcal{A}, \mathcal{P})$, $\boldsymbol{\pi}_{rt}$, $\boldsymbol{\pi}_t$, $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}_{\mathcal{I}}$, $\boldsymbol{\lambda}_{\mathcal{I}}$, $\boldsymbol{\mathcal{I}}_t$, $\boldsymbol{\alpha}_{\mathcal{S}}$, $\boldsymbol{\lambda}_{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_t$, $\boldsymbol{\lambda}_{\zeta}$, $\boldsymbol{\zeta}_t$, $\boldsymbol{\mathcal{I}}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\mathcal{I}}_t^{\mathbf{CM}}$, $\boldsymbol{\mathcal{I}}_t^{\mathbf{M}}$, $\boldsymbol{\mathcal{S}}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\mathcal{S}}_t^{\mathbf{CM}}$, $\boldsymbol{\mathcal{S}}_t^{\mathbf{M}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{UM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{CM}}$, $\boldsymbol{\lambda}_{\zeta}^{\mathbf{M}}$, $\boldsymbol{\zeta}_{rt}^{\mathbf{UM}}$, $\boldsymbol{\zeta}_t^{\mathbf{CM}}$, $\boldsymbol{\zeta}_t^{\mathbf{M}}\rangle$ *be an LGC-Com GRM with* (ζ_{tij2l}^{CM}, ζ_{tij2l}^{UM}, ζ_{tijkl}^{M})*-congeneric variables and strong conditional independence as defined by Definitions 5.1, 5.2 and 5.3. The parameters of the LGC-Com GRM with strong conditional independence and indicator-specific intercept, slope and state residual variables* $\mathcal{I}_{tij1}$, $\mathcal{S}_{tij1}$, $\mathcal{I}_{tij2}^{CM}$, $\mathcal{S}_{tij2}^{CM}$, $\mathcal{I}_{tijk}^{M}$, $\mathcal{S}_{tijk}^{M}$, $\mathcal{I}_{rtij2}^{UM}$, $\mathcal{S}_{rtij2}^{UM}$, *and* ζ_{tij1l} *are identified if*

1. *either one factor loading* $\lambda_{\zeta ij1l}$, $\lambda_{\zeta ij2l}^{CM}$, $\lambda_{\zeta ijkl}^{M}$, *and* $\lambda_{\zeta ij2l}^{UM}$ *for each factor* ζ_{tij1l}, ζ_{tj2l}^{CM}, ζ_{tjkl}^{M}, *and* ζ_{rtj2l}^{UM}, *or the variance of the factors is set to any real value larger than 0, and*
2. *one of the following conditions hold:*
   (a) $i_j = 2$, $j \geq 2$, $k \geq 2$, $l \geq 3$, *and* $\Phi_{\zeta B}$ *as well as* $\Phi_{\zeta W}$ *contain substantial permissible intercorrelations among the latent state residual as well as the latent state residual method variables,*
   (b) $i_j \geq 3$ *for all* j, $j \geq 1$, $k \geq 2$, $l \geq 3$, *and* $\Phi_{\zeta B}$ *contains substantial (permissible) intercorrelations between the latent state residuals* ζ_{tij1l},
   (c) $i_j \geq 3$ *for all* j, $j \geq 1$, $k \geq 3$, *and* $l \geq 3$,

   and
3. *either one threshold* κ_{sij11} *of the reference-method indicators on the first measurement occasion* Y_{tij11} *or the mean of the latent intercept factor* $\mathcal{I}_{tij1}$ *is set to any real value (e.g., zero) for all i and j, and*
4. *either one threshold* κ_{sijk1} *or* $\alpha_{\mathcal{I}ijk}$ *of the non-reference-method indicators on the first measurement occasion,* Y_{tijk1}, *is set to any real value (e.g., zero) for all i, j, and k, and*

Remarks. Theorem 5.8 states the conditions under which the parameters of the LGC-Com GRM are identified without further restrictions on loading, variance or threshold parameters than those that are necessary to assign a scale to the latent variables (and those that are imposed due to the prerequisite of invariant threshold parameters over time).

Condition (1) and (2) identify the parameters of the LGC-Com GRM covariance structure (that is, the parameters in the matrices $\Lambda_{\xi B}$, $\Lambda_{\xi W}$, $\Lambda_{\zeta B}$, $\Lambda_{\zeta W}$, $\Phi_{\xi B}$, $\Phi_{\xi W}$, $\Phi_{\zeta B}$, $\Phi_{\zeta W}$), given the polychoric correlations between the variables Y^*_{rtij2l} and Y^*_{tijkl} (or π_{rtij2l} and π_{tijkl}).

Note that conditions 2.(a) - 2.(c) in Theorem 5.8 define the necessary numbers of indicators, constructs, methods and measurement occasions needed to identify the LGC-Com GRM without imposing further assumptions. However, like the LST-Com GRM, the LGC-Com GRM can also be estimated for smaller designs when a few additional assumptions are imposed.

Conditions 3. and 4. are needed for the identification of the threshold variables κ_{sijkl}, intercept parameters $\alpha_{\mathcal{I}ijk}$, $\alpha_{\mathcal{S}ijk}$ as well as of the means of the latent response variables μ_{ijkl} and latent intercept and slope variables. Note that these parameters are defined with conditions 3. and 4., as invariance of the threshold parameters κ_{sijkl} over measurement is a prerequisite and assumed by definition of the LGC-Com GRM.

The LGC-Com GRM with common latent (method) intercept and slope factors for all indicators needs similar identifiability conditions, with the addition that for each construct j, $\delta_{\mathcal{I}ij1}$ and $\delta_{\mathcal{S}ij1}$ have to be set to any real value, preferably zero, for the same indicator i. The remaining $\delta_{\mathcal{I}ij1}$ could be estimated if the thresholds κ_{sij11} were set invariant over items i, however, for the ease of interpretation, it is recommended to set all $\delta_{\mathcal{I}ij1}$ to zero (and analogously for $l > 1$).

In the case were strong conditional independence does not hold, i.e., in the LGC-Com GRM with conditional independence as defined by Definition 5.2 additional identification conditions are necessary. For instance, in the most typical case of an AR(1) process (i.e., only regressions between adjacent measurement occasions are included), the model is identified with at least three measurement occasions and some additional assumptions on the loading parameters, regression parameters and state residual variances (Cole et al., 2005). For a more detailed treatment of autoregressive processes in LST and LGC models see, e.g., Cole et al. (2005), Eid et al. (2012), the online appendices of Bishop et al. (2015), or Hamaker (2005), Hamaker et al. (2015).

The identification of the parameters κ_{sijkl}, $\alpha_{\mathcal{I}ijk}$, $\alpha_{\mathcal{S}ijk}$, and the latent intercept and slope means follows, with only minor differences, the same lines as in the LS-Com GRM and is explained in detail in section 2.13. The identification of the parameters in the matrices $\Lambda_{\xi B}$, $\Lambda_{\xi W}$, $\Phi_{\xi B}$, and $\Phi_{\xi W}$ is derived in detail below. Given these parameters, the identification of the parameters in the matrices $\Lambda_{\zeta B}$, $\Lambda_{\zeta W}$, $\Phi_{\zeta B}$, and $\Phi_{\zeta W}$ is identical to the LST-Com GRM and described by Courvoisier (2006) and Koch (2013).

Proofs.

The following proofs concern the identification of the parameters in the matrices $\Lambda_{\xi B}$, $\Lambda_{\xi W}$, $\Phi_{\xi B}$, *and* $\Phi_{\xi W}$. *Parameters that are identified in previous identification steps will not be replaced by parameters of the observed variables. As a starting point for the identification, the covariances of the latent response variables are specified. These build on Theorem 5.5, that is, the correlations that are zero by definition are already excluded from the following equations. Furthermore, it is assumed that the covariance structure of the latent response variables* $\pi_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$ *(i.e., the expectations of the within-level latent response variables over clusters) is available. This is a standard assumption stating that cluster-level random effects are estimable in multilevel IRT models and is for instance described in Rabe-Hesketh et al. (2005) or, using Bayesian estimation with a Gibbs sampler, in Fox and Glas (2001). Covariances between the latent response variables are given by polychoric correlations of the observed ordinal variables and are estimated as described in Section 2.13.*

Identification is shown for an LGC-Com GRM with two methods, a structurally different self-report ($k = 1$) *and one set of interchangeable raters* ($k = 2$). *Identification of the variance, covariance and loading parameters for the variables* $\mathcal{I}^{M}_{tijk}$ *and* $\mathcal{S}^{M}_{tijk}$ *for a structurally different method* $k \geq 3$ *works exactly along the same lines as for the variables* $\mathcal{I}^{CM}_{tij2}$ *and* $\mathcal{S}^{CM}_{tij2}$ *as derived below.*

Denote $\pi_{tij2l} = S_{tij2l} = \mathbb{E}[\pi_{rtij2l} \mid p_T, p_{TS_l}]$. *Then, for* $l, l' > 1$, $l \neq l'$, $i \neq i'$, $k \neq 1$, *the covariances between the latent response variables on the between-level are given by*

$$\begin{aligned}
Cov(\pi_{tij11},\pi_{tij1l}) &= Var(\mathcal{I}_{tij1})+(l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) && (5.10.1)\\
Cov(\pi_{tij1l},\pi_{tij1l'}) &= Var(\mathcal{I}_{tij1})+(l-1)(l'-1)Var(\mathcal{S}_{tij1}) && (5.10.2)\\
&\quad +((l-1)+(l'-1))Cov(\mathcal{I}_{t111},\mathcal{S}_{t111})\\
Cov(\pi_{tij11},\pi_{ti'j1l}) &= Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+(l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1}) && (5.10.3)\\
Cov(\pi_{tij1l},\pi_{ti'j1l'}) &= Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+(l'-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1}) && (5.10.4)\\
&\quad +(l-1)Cov(\mathcal{S}_{tij1},\mathcal{I}_{ti'j1})+(l-1)(l'-1)Cov(\mathcal{S}_{tij1},\mathcal{S}_{ti'j1})
\end{aligned}$$

$$\begin{aligned}
Cov(\pi_{tij1l},\pi_{tij21}) &= \lambda_{\mathcal{I}ij2}Var(\mathcal{I}_{tij1})+(l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) && (5.10.5)\\
&\quad +(l-1)Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})\\
Cov(\pi_{tij11},\pi_{tij2l}) &= \lambda_{\mathcal{I}ij2}Var(\mathcal{I}_{tij1})+(l-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) && (5.10.6)\\
&\quad +(l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})\\
Cov(\pi_{tij1l},\pi_{tij2l'}) &= \lambda_{\mathcal{I}ij2}Var(\mathcal{I}_{tij1})+(l-1)(l'-1)\lambda_{\mathcal{S}ij2}Var(\mathcal{S}_{tij1}) && (5.10.7)\\
&\quad +(l-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})+(l'-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})\\
&\quad +(l-1)Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})+(l'-1)Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})
\end{aligned}$$

$$\begin{aligned}
Cov(\pi_{tij21},\pi_{tij2l}) &= (\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1})+Var(\mathcal{I}^{CM}_{tij2}) && (5.10.8)\\
&\quad +(l-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})+(l-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})\\
&\quad +(l-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})+(l-1)Cov(\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{tij2})\\
Cov(\pi_{tij2l},\pi_{tij2l'}) &= (\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1})+(l-1)(l'-1)(\lambda_{\mathcal{S}ij2})^2Var(\mathcal{S}_{tij1}) && (5.10.9)\\
&\quad +Var(\mathcal{I}^{CM}_{tij2})+(l-1)(l'-1)Var(\mathcal{S}^{CM}_{tij2})\\
&\quad +((l-1)+(l'-1))\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})\\
&\quad +((l-1)+(l'-1))\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})\\
&\quad +((l-1)+(l'-1))\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})\\
&\quad +((l-1)+(l'-1))Cov(\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{tij2})\\
Cov(\pi_{tij11},\pi_{ti'j2l}) &= \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+Cov(\mathcal{I}_{tij1},\mathcal{I}^{CM}_{ti'j2}) && (5.10.10)\\
&\quad +(l-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1})+(l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{ti'j2})
\end{aligned}$$

$$\begin{aligned}
Cov(\pi_{tij1l},\pi_{ti'j21}) &= \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+Cov(\mathcal{I}_{tij1},\mathcal{I}^{CM}_{ti'j2}) && (5.10.11)\\
&\quad +(l-1)\lambda_{\mathcal{I}i'j2}Cov(\mathcal{S}_{tij1},\mathcal{I}_{ti'j1})+(l-1)Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{ti'j2})\\
Cov(\pi_{tij1l},\pi_{ti'j2l'}) &= \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+Cov(\mathcal{I}_{tij1},\mathcal{I}^{CM}_{ti'j2}) && (5.10.12)\\
&\quad +(l'-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1})+(l'-1)Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{ti'j2})\\
&\quad +(l-1)\lambda_{\mathcal{I}i'j2}Cov(\mathcal{S}_{tij1},\mathcal{I}_{ti'j1})+(l-1)Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{ti'j2})\\
&\quad +(l-1)(l'-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{S}_{tij1},\mathcal{S}_{ti'j1})+(l-1)(l'-1)Cov(\mathcal{S}_{tij1},\mathcal{S}^{CM}_{ti'j2})\\
Cov(\pi_{tij21},\pi_{ti'j2l}) &= \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1})+\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{I}^{CM}_{ti'j2}) && (5.10.13)\\
&\quad +\lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{ti'j1},\mathcal{I}^{CM}_{tij2})+(l-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1})\\
&\quad +(l-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{ti'j2})+(l-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{S}_{ti'j1},\mathcal{I}^{CM}_{tij2})\\
&\quad +Cov(\mathcal{I}^{CM}_{tij2},\mathcal{I}^{CM}_{ti'j2})+(l-1)Cov(\mathcal{I}^{CM}_{tij1},\mathcal{S}^{CM}_{ti'j2})
\end{aligned}$$

$$
\begin{aligned}
Cov(\pi_{tij2l},\pi_{tij'2l'}) = {} & \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{I}_{ti'j1}) \\
& + (l-1)(l'-1)\lambda_{\mathcal{S}ij2}\lambda_{\mathcal{S}i'j2}Cov(\mathcal{S}_{tij1},\mathcal{S}_{ti'j1}) \\
& + \lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{I}^{CM}_{ti'j2}) + \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{ti'j1},\mathcal{I}^{CM}_{tij2}) \\
& + (l'-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{ti'j1}) + (l-1)\lambda_{\mathcal{I}i'j2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{ti'j1},\mathcal{S}_{tij1}) \\
& + (l'-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{ti'j2}) + (l-1)\lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{ti'j1},\mathcal{S}^{CM}_{tij2}) \\
& + (l-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{ti'j2}) + (l'-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{S}_{ti'j1},\mathcal{I}^{CM}_{tij2}) \\
& + (l-1)(l'-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{S}^{CM}_{ti'j2}) + (l-1)(l'-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{S}_{ti'j1},\mathcal{S}^{CM}_{tij2}) \\
& + (l-1)Cov(\mathcal{I}^{CM}_{ti'j2},\mathcal{S}^{CM}_{tij2}) + (l'-1)Cov(\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{ti'j2}) \\
& + Cov(\mathcal{I}^{CM}_{tij2},\mathcal{I}^{CM}_{ti'j2}) + (l-1)(l'-1)Cov(\mathcal{S}^{CM}_{tij2},\mathcal{S}^{CM}_{ti'j2})
\end{aligned} \tag{5.10.14}
$$

And, for $l,l' > 1$, $l \neq l'$, $i \neq i'$, $k \neq 1$, the covariances between the latent response variables on the within-level are given by

$$Cov(\pi_{rtij21},\pi_{rtij2l}) = Cov(\pi_{tij21},\pi_{tij2l}) + Var(\mathcal{I}^{UM}_{rtij2}) + (l-1)Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{S}^{UM}_{rtij2}) \tag{5.10.15}$$

$$
\begin{aligned}
Cov(\pi_{rtij2l},\pi_{rtij2l'}) = {} & Cov(\pi_{tij2l},\pi_{tij2l'}) + Var(\mathcal{I}^{UM}_{rtij2}) + (l-1)(l'-1)Var(\mathcal{S}^{UM}_{rtij2}) \\
& + ((l-1)+(l'-1))Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{S}^{UM}_{rtij2})
\end{aligned} \tag{5.10.16}
$$

$$Cov(\pi_{rtij21},\pi_{rti'j2l}) = Cov(\pi_{tij21},\pi_{ti'j2l}) + Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{I}^{UM}_{rti'j2}) + (l-1)Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{S}^{UM}_{rti'j2}) \tag{5.10.17}$$

$$
\begin{aligned}
Cov(\pi_{rtij2l},\pi_{rti'j2l'}) = {} & Cov(\pi_{tij2l},\pi_{ti'j2l'}) + Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{I}^{UM}_{rti'j2}) + (l'-1)Cov(\mathcal{I}^{UM}_{rtij2},\mathcal{S}^{UM}_{rti'j2}) \\
& + (l-1)Cov(\mathcal{I}^{UM}_{rti'j1},\mathcal{S}^{UM}_{rtij2}) + (l-1)(l'-1)Cov(\mathcal{S}^{UM}_{rtij2},\mathcal{S}^{UM}_{rti'j2})
\end{aligned} \tag{5.10.18}
$$

Identification of the parameters in $\Phi_{\xi B}$ and $\Lambda_{\xi B}$.

Identification of $Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})$
By Equation (5.10.1) it holds that

$$Var(\mathcal{I}_{tij1}) = Cov(\pi_{tij11},\pi_{tij1l}) - (l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) \tag{5.10.19}$$

$$Cov(\pi_{tij11},\pi_{tij1l'}) = Var(\mathcal{I}_{tij1}) + (l'-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) \tag{5.10.20}$$

Inserting 5.10.19 into 5.10.20 yields

$$
\begin{aligned}
Cov(\pi_{tij11},\pi_{tij1l'}) &= Cov(\pi_{tij11},\pi_{tij1l}) - (l-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) + (l'-1)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) \\
&= Cov(\pi_{tij11},\pi_{tij1l}) + (l'-l)Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})
\end{aligned}
$$

That is, $Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})$ is identified as:

$$Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) = \frac{1}{(l'-l)}\big(Cov(\pi_{tij11},\pi_{tij1l'}) - Cov(\pi_{tij11},\pi_{tij1l})\big)$$

Identification of $Var(\mathcal{I}_{tij1})$
Inserting $Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})$ into Equation (5.10.19) identifies $Var(\mathcal{I}_{tij1})$ as

$$Var(\mathcal{I}_{tij1}) = Cov(\pi_{tij11}, \pi_{tij1l}) - \frac{(l-1)}{(l'-l)}\big(Cov(\pi_{tij11}, \pi_{tij1l'}) - Cov(\pi_{tij11}, \pi_{tij1l})\big)$$

Identification of $Var(\mathcal{S}_{tij1})$
By Equation (5.10.2) *it holds that*

$$(l-1)(l'-1)Var(\mathcal{S}_{tij1}) = Cov(\pi_{tij1l}, \pi_{tij1l'}) - Var(\mathcal{I}_{tij1}) - ((l-1)+(l'-1))Cov(\mathcal{I}_{t111}, \mathcal{S}_{t111})$$

That is, $Var(\mathcal{S}_{tij1})$ is identified as,

$$Var(\mathcal{S}_{tij1}) = \frac{Cov(\pi_{tij1l}, \pi_{tij1l'}) - Var(\mathcal{I}_{tij1}) - ((l-1)+(l'-1))Cov(\mathcal{I}_{t111}, \mathcal{S}_{t111})}{(l-1)(l'-1)}$$

Identification of $Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1})$
By Equation (5.10.3) *it holds that*

$$Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) = Cov(\pi_{tij11}, \pi_{ti'j1l}) - (l-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \tag{5.10.21}$$

$$Cov(\pi_{tij11}, \pi_{ti'j1l'}) = Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) + (l'-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \tag{5.10.22}$$

Inserting Equation (5.10.21) *into Equation* (5.10.22) *yields*

$$\begin{aligned} Cov(\pi_{tij11}, \pi_{ti'j1l'}) &= Cov(\pi_{tij11}, \pi_{ti'j1l}) - (l-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) + (l'-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \\ &= Cov(\pi_{tij11}, \pi_{ti'j1l}) + (l'-l)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \end{aligned}$$

That is, $Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1})$ is identified as:

$$Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) = \frac{1}{(l'-l)}\big(Cov(\pi_{tij11}, \pi_{ti'j1l'}) - Cov(\pi_{tij11}, \pi_{ti'j1l})\big)$$

Identification of $Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1})$
Inserting $Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1})$ into Equation (5.10.21) *identifies $Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1})$ as*

$$Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) = Cov(\pi_{tij11}, \pi_{ti'j1l}) - \frac{(l-1)}{(l'-l)}\big(Cov(\pi_{tij11}, \pi_{ti'j1l'}) - Cov(\pi_{tij11}, \pi_{ti'j1l})\big)$$

Identification of $Cov(\mathcal{S}_{tij1}, \mathcal{S}_{ti'j1})$
Then, by Equation (5.10.4), *$Cov(\mathcal{S}_{tij1}, \mathcal{S}_{ti'j1})$ is identified by*

$$Cov(\mathcal{S}_{tij1}, \mathcal{S}_{ti'j1}) = \frac{Cov(\pi_{tij1l}, \pi_{ti'j1l'}) - Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) - (l'-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) - (l-1)Cov(\mathcal{S}_{tij1}, \mathcal{I}_{ti'j1})}{(l-1)(l'-1)}$$

Identification of $\lambda_{\mathcal{I}ij2}$

By Equation (5.10.5) *it holds that*

$$Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) = \frac{Cov(\pi_{tij1l}, \pi_{tij21}) - \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) - (l-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1})}{(l-1)} \tag{5.10.23}$$

$$Cov(\pi_{tij1l'}, \pi_{tij21}) = \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) + (l'-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) + (l'-1) Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) \tag{5.10.24}$$

Inserting 5.10.23 into 5.10.24 yields

$$\begin{aligned} Cov(\pi_{tij1l'}, \pi_{tij21}) &= \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) + (l'-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ &+ \frac{(l'-1)}{(l-1)} \big(Cov(\pi_{tij1l}, \pi_{tij21}) - \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) - (l-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \big) \\ &= \frac{(l-l')}{(l-1)} \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) + \frac{(l'-1)}{(l-1)} Cov(\pi_{tij1l}, \pi_{tij21}) \end{aligned}$$

That is, $\lambda_{\mathcal{I}ij2}$ is identified as:

$$\lambda_{\mathcal{I}ij2} = \frac{(l-1)}{(l-l')} \frac{Cov(\pi_{tij1l'}, \pi_{tij21})}{Var(\mathcal{I}_{tij1})} - \frac{(l'-1)}{(l-l')} \frac{Cov(\pi_{tij1l}, \pi_{tij21})}{Var(\mathcal{I}_{tij1})}$$

Identification of $Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2})$

Inserting $\lambda_{\mathcal{I}ij2}$ into Equation (5.10.5) *identifies $Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij1})$ by*

$$Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) = \frac{Cov(\pi_{tij1l}, \pi_{tij21})}{(l-1)} - \frac{\lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1})}{(l-1)} - Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1})$$

Identification of $\lambda_{\mathcal{S}ij2}$

By Equation (5.10.7) *it holds that*

$$\begin{aligned} Cov(\pi_{tij1l}, \pi_{tij2l'}) &= \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) + (l-1)(l'-1) \lambda_{\mathcal{S}ij2} Var(\mathcal{S}_{tij1}) \\ &+ (l-1) \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) + (l'-1) \lambda_{\mathcal{S}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ &+ (l-1) Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) + (l'-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{tij2}) \end{aligned} \tag{5.10.25}$$

$$\begin{aligned} Cov(\pi_{tij1l'}, \pi_{tij2l}) &= \lambda_{\mathcal{I}ij2} Var(\mathcal{I}_{tij1}) + (l-1)(l'-1) \lambda_{\mathcal{S}ij2} Var(\mathcal{S}_{tij1}) \\ &+ (l'-1) \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) + (l-1) \lambda_{\mathcal{S}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ &+ (l'-1) Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) + (l-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{tij2}) \end{aligned} \tag{5.10.26}$$

Subtracting (5.10.26) *from* (5.10.25) *yields*

$$\begin{aligned} Cov(\pi_{tij1l}, \pi_{tij2l'}) - Cov(\pi_{tij1l'}, \pi_{tij2l}) &= (l-l') \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) + (l'-l) \lambda_{\mathcal{S}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) \\ &+ (l-l') Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2}) + (l'-l) Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{tij2}) \end{aligned}$$

and hence

$$Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2}) = \frac{Cov(\pi_{tij1l},\pi_{tij2l'})}{(l'-l)} - \frac{Cov(\pi_{tij1l'},\pi_{tij2l})}{(l'-l)}$$
$$+\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) - \lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) + Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})$$

Inserting this into (5.10.26) *yields*

$$Cov(\pi_{tij1l'},\pi_{tij2l}) = \lambda_{\mathcal{I}ij2}Var(\mathcal{I}_{tij1}) + (l-1)(l'-1)\lambda_{\mathcal{S}ij2}Var(\mathcal{S}_{tij1})$$
$$+((l'-1)+(l-1))\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})$$
$$+((l'-1)+(l-1))Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2}) + \frac{(l-1)}{(l'-l)}Cov(\pi_{tij1l},\pi_{tij2l'}) - \frac{(l-1)}{(l'-l)}Cov(\pi_{tij1l'},\pi_{tij2l})$$

and hence $\lambda_{\mathcal{S}ij2}$ *is identified by*

$$\lambda_{\mathcal{S}ij2} = \frac{1}{(l-1)(l'-1)Var(\mathcal{S}_{tij1})}\Big(-\lambda_{\mathcal{I}ij2}Var(\mathcal{I}_{tij1})$$
$$+\frac{(l'-1)}{(l'-l)}Cov(\pi_{tij1l'},\pi_{tij2l}) - \frac{(l-1)}{(l'-l)}Cov(\pi_{tij1l},\pi_{tij2l'})$$
$$-((l'-1)+(l-1))\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) - ((l'-1)+(l-1))Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})\Big)$$

Identification of $Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})$

Then, by Equation (5.10.6), $Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})$ *is identified by*

$$Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2}) = \frac{1}{(l-1)}Cov(\pi_{tij11},\pi_{tij2l}) - \frac{\lambda_{\mathcal{I}ij2}}{(l-1)}Var(\mathcal{I}_{tij1}) - \lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1})$$

Identification of $Var(\mathcal{I}^{CM}_{tij2})$

By Equation (5.10.8) *it holds that*

$$Cov(\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{tij2}) = \frac{1}{(l-1)}Cov(\pi_{tij21},\pi_{tij2l}) - \frac{1}{(l-1)}(\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1}) \quad (5.10.27)$$
$$-\frac{1}{(l-1)}Var(\mathcal{I}^{CM}_{tij2}) - \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) - \lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})$$
$$-\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2})$$
$$Cov(\pi_{tij21},\pi_{tij2l'}) = (\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1}) + Var(\mathcal{I}^{CM}_{tij2}) \quad (5.10.28)$$
$$+(l'-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}_{tij1}) + (l'-1)\lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1},\mathcal{S}^{CM}_{tij2})$$
$$+(l'-1)\lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1},\mathcal{I}^{CM}_{tij2}) + (l'-1)Cov(\mathcal{I}^{CM}_{tij2},\mathcal{S}^{CM}_{tij2})$$

Inserting (5.10.27) *into* (5.10.28) *yields*

$$Cov(\pi_{tij21},\pi_{tij2l'}) = \frac{(l-l')}{(l-1)}(\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1}) + \frac{(l-l')}{(l-1)}Var(\mathcal{I}^{CM}_{tij2}) + \frac{(l'-1)}{(l-1)}Cov(\pi_{tij21},\pi_{tij2l})$$

Hence, $Var(\mathcal{I}^{CM}_{tij2})$ *is identified by*

$$Var(\mathcal{I}^{CM}_{tij2}) = \frac{(l-1)}{(l-l')}Cov(\pi_{tij21},\pi_{tij2l'}) - (\lambda_{\mathcal{I}ij2})^2Var(\mathcal{I}_{tij1}) - Cov(\pi_{tij21},\pi_{tij2l})$$

Identification of $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{tij2})$

Then, $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{tij2})$ *is identified by Equation* (5.10.8)*:*

$$Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{tij2}) = \frac{1}{(l-1)}Cov(\pi_{tij21}, \pi_{tij2l}) - \frac{1}{(l-1)}(\lambda_{\mathcal{I}ij2})^2 Var(\mathcal{I}_{tij1}) - \frac{1}{(l-1)}Var(\mathcal{I}^{CM}_{tij2})$$
$$- \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}ij2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{tij1}) - \lambda_{\mathcal{I}ij2}Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{tij2}) - \lambda_{\mathcal{S}ij2}Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{tij2})$$

Identification of $Var(\mathcal{S}^{CM}_{tij2})$

Then $Var(\mathcal{S}^{CM}_{tij2})$ *is identified by rearranging Equation* (5.10.9).

Identification of $Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2})$

By Equation (5.10.10) *it holds that*

$$Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) = Cov(\pi_{tij11}, \pi_{ti'j2l}) - \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1})$$
$$- (l-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) - (l-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) \quad (5.10.29)$$

$$Cov(\pi_{tij11}, \pi_{ti'j2l'}) = \lambda_{\mathcal{I}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) + Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2})$$
$$+ (l'-1)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) + (l'-1)Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) \quad (5.10.30)$$

Inserting Equation 5.10.29 into Equation 5.10.30 yields

$$Cov(\pi_{tij11}, \pi_{ti'j2l'}) = Cov(\pi_{tij11}, \pi_{ti'j2l})$$
$$+ (l'-l)\lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) + (l'-l)Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2})$$

That is, $Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2})$ *is identified as*

$$Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) = \frac{1}{(l'-l)}Cov(\pi_{tij11}, \pi_{ti'j2l'}) - \frac{1}{(l'-l)}Cov(\pi_{tij11}, \pi_{ti'j2l})$$
$$- \lambda_{\mathcal{S}i'j2}Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1})$$

Identification of $Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2})$

Then, $Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2})$ *is identified by Equation* (5.10.29).

Identification of $Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{ti'j2})$

Then, $Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{ti'j2})$ *is identified by rearranging Equation* (5.10.11)*:*

$$Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) = \frac{1}{(l-1)}Cov(\pi_{tij1l}, \pi_{ti'j21}) - \frac{\lambda_{\mathcal{I}i'j2}}{(l-1)}Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1})$$
$$- \frac{1}{(l-1)}Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) - \lambda_{\mathcal{I}i'j2}Cov(\mathcal{S}_{tij1}, \mathcal{I}_{ti'j1})$$

Identification of $Cov(\mathcal{S}_{tij1}, \mathcal{S}^{CM}_{ti'j2})$

Then, $Cov(\mathcal{S}_{tij1}, \mathcal{S}^{CM}_{ti'j2})$ is identified by rearranging Equation (5.10.12)*:*

$$\begin{aligned}
Cov(\mathcal{S}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) = \frac{1}{(l-1)(l'-1)} \Big(& Cov(\pi_{tij1l}, \pi_{ti'j2l'}) - \lambda_{\mathcal{I}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) - Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) \\
& - (l'-1)\lambda_{\mathcal{S}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) - (l'-1) Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) \\
& - (l-1)\lambda_{\mathcal{I}i'j2} Cov(\mathcal{S}_{tij1}, \mathcal{I}_{ti'j1}) - (l-1) Cov(\mathcal{S}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) \\
& - (l-1)(l'-1)\lambda_{\mathcal{S}i'j2} Cov(\mathcal{S}_{tij1}, \mathcal{S}_{ti'j1}) \Big)
\end{aligned}$$

Identification of $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{ti'j2})$

By Equation (5.10.13) *it holds that*

$$\begin{aligned}
Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{I}^{CM}_{ti'j2}) = {} & Cov(\pi_{tij21}, \pi_{ti'j2l}) - \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{I}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) \\
& - \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) - \lambda_{\mathcal{I}i'j2} Cov(\mathcal{I}_{ti'j1}, \mathcal{I}^{CM}_{tij2}) \\
& - (l-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) - (l-1)\lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) \\
& - (l-1)\lambda_{\mathcal{S}i'j2} Cov(\mathcal{S}_{ti'j1}, \mathcal{I}^{CM}_{tij2}) - (l-1) Cov(\mathcal{I}^{CM}_{tij1}, \mathcal{S}^{CM}_{ti'j2})
\end{aligned} \tag{5.10.31}$$

$$\begin{aligned}
Cov(\pi_{tij21}, \pi_{ti'j2l'}) = {} & \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{I}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{I}_{ti'j1}) + \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{I}^{CM}_{ti'j2}) \\
& + \lambda_{\mathcal{I}i'j2} Cov(\mathcal{I}_{ti'j1}, \mathcal{I}^{CM}_{tij2}) + (l'-1)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \\
& + (l'-1)\lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) + (l'-1)\lambda_{\mathcal{S}i'j2} Cov(\mathcal{S}_{ti'j1}, \mathcal{I}^{CM}_{tij2}) \\
& + Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{I}^{CM}_{ti'j2}) + (l'-1) Cov(\mathcal{I}^{CM}_{tij1}, \mathcal{S}^{CM}_{ti'j2})
\end{aligned} \tag{5.10.32}$$

Inserting Equation 5.10.31 into Equation 5.10.32 yields

$$\begin{aligned}
Cov(\pi_{tij21}, \pi_{ti'j2l'}) = {} & Cov(\pi_{tij21}, \pi_{ti'j2l}) + (l'-l)\lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \\
& + (l'-l)\lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) + (l'-l)\lambda_{\mathcal{S}i'j2} Cov(\mathcal{S}_{ti'j1}, \mathcal{I}^{CM}_{tij2}) \\
& + (l'-l) Cov(\mathcal{I}^{CM}_{tij1}, \mathcal{S}^{CM}_{ti'j2})
\end{aligned}$$

That is, $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{S}^{CM}_{ti'j2})$ is identified as

$$\begin{aligned}
Cov(\mathcal{I}^{CM}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) = {} & \frac{1}{(l'-l)} Cov(\pi_{tij21}, \pi_{ti'j2l'}) - \frac{1}{(l'-l)} Cov(\pi_{tij21}, \pi_{ti'j2l}) - \lambda_{\mathcal{I}ij2}\lambda_{\mathcal{S}i'j2} Cov(\mathcal{I}_{tij1}, \mathcal{S}_{ti'j1}) \\
& - \lambda_{\mathcal{I}ij2} Cov(\mathcal{I}_{tij1}, \mathcal{S}^{CM}_{ti'j2}) - \lambda_{\mathcal{S}i'j2} Cov(\mathcal{S}_{ti'j1}, \mathcal{I}^{CM}_{tij2})
\end{aligned}$$

Identification of $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{I}^{CM}_{ti'j2})$

Then $Cov(\mathcal{I}^{CM}_{tij2}, \mathcal{I}^{CM}_{ti'j2})$ is identified by Equation (5.10.31).

Identification of $Cov(\mathcal{S}^{CM}_{tij2}, \mathcal{S}^{CM}_{ti'j2})$

Then $Cov(\mathcal{S}^{CM}_{tij2}, \mathcal{S}^{CM}_{ti'j2})$ is identified by rearranging Equation (5.10.14).

Identification of the parameters in $\Phi_{\xi W}$ and $\Lambda_{\xi W}$.

Identification of $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2})$

By Equation (5.10.15) *it holds that*

$$Var(\mathcal{I}^{UM}_{tij2}) = Cov(\pi_{rtij21}, \pi_{rtij2l}) - Cov(\pi_{tij21}, \pi_{tij2l}) - (l-1)Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2}) \quad (5.10.33)$$

$$Cov(\pi_{rtij21}, \pi_{rtij2l'}) = Cov(\pi_{tij21}, \pi_{tij2l'}) + Var(\mathcal{I}^{UM}_{tij2}) + (l'-1)Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2}) \quad (5.10.34)$$

Inserting 5.10.33 into 5.10.34 yields

$$\begin{aligned} Cov(\pi_{rtij21}, \pi_{rtij2l'}) = {} & Cov(\pi_{tij21}, \pi_{tij2l'}) + Cov(\pi_{rtij21}, \pi_{rtij2l}) - Cov(\pi_{tij21}, \pi_{tij2l}) \\ & + (l'-l)Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2}) \end{aligned}$$

That is, $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2})$ is identified as:

$$Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2}) = \frac{1}{(l'-l)}\big(Cov(\pi_{rtij21}, \pi_{rtij2l'}) - Cov(\pi_{rtij21}, \pi_{rtij2l}) + Cov(\pi_{tij21}, \pi_{tij2l}) - Cov(\pi_{tij21}, \pi_{tij2l'})\big)$$

Identification of $Var(\mathcal{I}^{UM}_{tij2})$

Inserting $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{tij2})$ into Equation (5.10.33) *identifies $Var(\mathcal{I}^{UM}_{tij2})$.*

Identification of $Var(\mathcal{S}^{UM}_{tij2})$

By Equation (5.10.16) *it holds that*

$$\begin{aligned} (l-1)(l'-1)Var(\mathcal{S}^{UM}_{rtij2}) = {} & Cov(\pi_{rtij2l}, \pi_{rtij2l'}) - Cov(\pi_{tij2l}, \pi_{tij2l'}) - \\ & Var(\mathcal{I}^{UM}_{rtij2}) - ((l-1)+(l'-1))Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{S}^{UM}_{rtij2}) \end{aligned}$$

That is, $Var(\mathcal{S}^{UM}_{tij2})$ is identified as,

$$Var(\mathcal{S}^{UM}_{tij2}) = \frac{Cov(\pi_{rtij2l}, \pi_{rtij2l'}) - Cov(\pi_{tij2l}, \pi_{tij2l'}) - Var(\mathcal{I}^{UM}_{rtij2}) - ((l-1)+(l'-1))Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{S}^{UM}_{rtij2})}{(l-1)(l'-1)}$$

Identification of $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1})$

By Equation (5.10.17) *it holds that*

$$Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{I}^{UM}_{rti'j2}) = Cov(\pi_{rtij21}, \pi_{rti'j2l}) - Cov(\pi_{tij21}, \pi_{ti'j2l}) - (l-1)Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{S}^{UM}_{rti'j2}) \quad (5.10.35)$$

$$Cov(\pi_{rtij21}, \pi_{rti'j2l'}) = Cov(\pi_{tij21}, \pi_{ti'j2l'}) + Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{I}^{UM}_{rti'j2}) + (l'-1)Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{S}^{UM}_{rti'j2}) \quad (5.10.36)$$

Inserting Equation (5.10.35) *into Equation* (5.10.36) *yields*

$$Cov(\pi_{rtij21}, \pi_{rti'j2l'}) = Cov(\pi_{tij21}, \pi_{ti'j2l'}) + Cov(\pi_{rtij21}, \pi_{rti'j2l}) - Cov(\pi_{tij21}, \pi_{ti'j2l}) + (l'-l)Cov(\mathcal{I}^{UM}_{rtij2}, \mathcal{S}^{UM}_{rti'j2})$$

That is, $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1})$ is identified as:

$$Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1}) = \frac{1}{(l'-l)}\big(Cov(\pi_{rtij21}, \pi_{rti'j2l'}) - Cov(\pi_{tij21}, \pi_{ti'j2l'}) - Cov(\pi_{rtij21}, \pi_{rti'j2l}) + Cov(\pi_{tij21}, \pi_{ti'j2l})\big)$$

Identification of $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{I}^{UM}_{ti'j1})$

Inserting $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1})$ into Equation (5.10.35) *identifies $Cov(\mathcal{I}^{UM}_{tij2}, \mathcal{I}^{UM}_{ti'j1})$.*

Identification of $Cov(\mathcal{S}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1})$

Then, $Cov(\mathcal{S}^{UM}_{tij2}, \mathcal{S}^{UM}_{ti'j1})$ is identified by rearranging Equation (5.10.18).

Given the identification of the parameters in $\Lambda_{\xi B}$, $\Lambda_{\xi W}$, $\Phi_{\xi B}$, and $\Phi_{\xi W}$, the identification of the parameters in the matrices $\Lambda_{\zeta B}$, $\Lambda_{\zeta W}$, $\Phi_{\zeta B}$, and $\Phi_{\zeta W}$ is identical to the LST-Com GRM and described by Courvoisier (2006) and Koch (2013). For the identification of the mean structure and threshold parameters see Section 2.13 and 5.9.

Chapter 6

Analyzing MTMM Data with Bayesian Methods

6.1 Short introduction to Bayesian statistics

The simulation studies and data application discussed in the following chapters (7 and 8) make use of Bayesian estimation techniques. As the description of the methods and results build on theoretical terms from the field of Bayesian statistics and Markov chain Monte Carlo (MCMC) methods, the basic ideas and concepts of these methods shall be shortly introduced.

In the Bayesian estimation approach parameters are regarded as random variables with a probability distribution. In this framework, probability is considered to be a subjective belief rather than a long-run frequency, the way probability is mostly conceptualized in classical statistics (de Finetti, 1974; Jackman, 2009). Hence, while in classical statistics data are considered to be random and parameters to be fixed, in Bayesian statistics the data are considered to be fixed (once sampled) and parameters to be random entities, which are subject to uncertainty.

Bayesian statistics rely on using data to update prior assumptions regarding a relevant entity θ (e.g., a parameter), by applying Bayes theorem to probability distributions (Jackman, 2009). Let $\boldsymbol{\theta}$ be a parameter vector of interest, $\boldsymbol{y}$ a data vector, and let $p(\cdot)$ denote a probability density function. Making use of Bayes theorem, prior assumptions about $\boldsymbol{\theta}$, expressed in the prior probability $p(\boldsymbol{\theta})$, are updated using observed data, expressed in the likelihood of the data given $\boldsymbol{\theta}$, $p(\boldsymbol{y} \mid \boldsymbol{\theta}) = L(\boldsymbol{\theta}, \boldsymbol{y})$, by

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\boldsymbol{y})} = \frac{L(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta})}{\int p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{6.1}$$

The resulting posterior probability $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ represents the updated assumptions about $\boldsymbol{\theta}$ after taking the data into account. Hence, the posterior distribution can be understood as a compromise between the prior assumptions expressed in $p(\boldsymbol{\theta})$ and the information provided by the data. In this way, Bayesian estimation can be truly accumulative, with the possibility to explicitly integrate previous research results into the data analysis via the prior probability $p(\boldsymbol{\theta})$. That is, if informative priors are based on previous studies, the posterior distribution integrates different sources of information and is thereby more precise than the likelihood or the prior alone (Jackman, 2009).

In contrast to classical statistics, Bayesian data analysis does not rely on asymptotic arguments, making it preferable for the analysis of small samples. Estimation techniques that rely on large-sample properties, such as ML, tend to produce unreliable and unstable results in small samples (Asparouhov & Muthén, 2010b; Hox & Maas, 2001; Meuleman & Billiet, 2009), while Bayesian estimation has been found to produce more reliable results than ML for small samples (Song & Lee, 2012; Lee & Song, 2004) or in multilevel models with few observations on the between-level (Hox et al., 2012; Asparouhov & Muthén, 2010b; Baldwin & Fellingham, 2013). Additionally, the possibility to incorporate informative priors in the estimation process might facilitate estimation and further increase the applicability of complex models in small samples (Depaoli & Clifton, 2015; Holtmann et al., 2016; Lee et al., 2010).

However, with increasing sample sizes, the impact of the prior on the posterior distribution decreases, and asymptotically the posterior approximates the likelihood (Lynch, 2007; Song & Lee, 2012). This is the case as the likelihood depends on the sample size, such that the weight of the likelihood in applying Bayes theorem increases with increasing sample size. The prior, in contrast, does not depend on the sample size. The weight of the prior information in the updating process via Bayes Theorem is determined by the priors' informativeness. The informativeness of the prior is controlled by the prior variance, i.e., the variance of the prior distribution specified by the researcher.

Bayesian estimation techniques make it possible to analyze models that are computationally heavy or impossible to estimate with classical estimation techniques such as ML or WLSMV estimation (Asparouhov & Muthén, 2012; B. Muthén, 2010; Asparouhov & Muthén, 2010b). This concerns, for instance, models with categorical indicators and many latent variables. ML estimation of these models requires numerical integration with many dimensions of integration (B. Muthén, 2010), making the estimation process computationally demanding.

Another advantage of Bayesian analysis is that results with improper solutions and inadmissible parameter estimates can be avoided by assigning zero prior probability to these parameter spaces (Depaoli & Clifton, 2015; Hox et al., 2012). Furthermore, Bayesian methods allow researchers to compute credibility intervals for key quantities with unknown, potentially skewed distributions. Classical calculations of confidence intervals based on normal theory may be unreliable for these types of parameters (e.g., correlations). This feature is especially useful in MTMM analyses, where coefficients of interest are often expressed in terms of variance components, such as the consistency and method specificity coefficients.

The use of Bayesian data analysis and the number of available statistical software packages has grown immensely in the last decades, triggered by increasing computational power and the availability of MCMC methods.

6.2 Markov chain Monte Carlo methods

Bayesian analysis relies on sampling from complex distributions using MCMC methods, an iterative sampling process generating a Markov chain of draws from the posterior distribution. MCMC methods use the fact that the integral in the denominator of (6.1), which cannot be solved analytically, does not contain information about $\boldsymbol{\theta}$. That is,

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{L(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta})}{\int p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} = c^{-1} L(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}) \tag{6.2.1}$$

with

$$c = \int p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{6.2.2}$$

where the normalizing constant c guarantees that the posterior distribution is a probability density distribution. As c does not contain information about $\boldsymbol{\theta}$, it is not needed for MCMC sampling methods and estimation relies on the famous proportionality

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto L(\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}) \tag{6.2.3}$$

MCMC methods repeatedly draw samples from $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, which are then used to approximate the posterior distribution and compute summary statistics which can serve as point estimates. For instance, the 2.5% and the 97.5% quantiles of the posterior distribution serve to estimate a 95% credibility interval (CI) for the parameter estimate.
As $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ is unknown, generating independent draws from it is not possible. Instead, a Markov chain of draws is generated from $p(\boldsymbol{\theta} \mid \boldsymbol{y})$, which, in the limit, converges against the target distribution given a number of conditions are met (see, e.g., Jackman, 2009, chapter 4, for details).
MCMC algorithms and their implementations vary across software packages for Bayesian data analysis. Mplus (L. K. Muthén & Muthén, 1998-2012) was chosen for the estimation of the models in the following simulation studies and applications as it is one of the most widely applied programs for structural equation modeling and due to its comparatively fast estimation.
Mplus uses an MCMC algorithm that is based on the Gibbs sampler. Gibbs sampling (Geman & Geman, 1984) is an MCMC algorithm that allows to sample from multivariate distributions $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_p)$ if the full conditional distributions $p(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_{-i})$, where $\boldsymbol{\theta}_{-i}$ denotes $\boldsymbol{\theta}$ without $\boldsymbol{\theta}_i$, are known. Suppose the parameter vector $\boldsymbol{\theta}$ is partitioned into p blocks or sub-vectors $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)'$. For b+R iterations, the Gibbs sampler then performs the following steps (Jackman, 2009; Gelman et al., 2014):

- For r in 1 to (b+R)

  1) draw $\boldsymbol{\theta}_1^{(r)}$ from $p(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(r-1)}, \ldots, \boldsymbol{\theta}_p^{(r-1)}, \boldsymbol{y})$

  2) draw $\boldsymbol{\theta}_2^{(r)}$ from $p(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^{(r)}, \boldsymbol{\theta}_3^{(r-1)}, \ldots, \boldsymbol{\theta}_p^{(r-1)}, \boldsymbol{y})$

     $\vdots$

  p) draw $\boldsymbol{\theta}_p^{(r)}$ from $p(\boldsymbol{\theta}_p \mid \boldsymbol{\theta}_1^{(r)}, \boldsymbol{\theta}_2^{(r)}, \ldots, \boldsymbol{\theta}_{p-1}^{(r)}, \boldsymbol{y})$

- end for
- Discard first b draws as burn in

That is, the Gibbs sampler cycles through the different components of $\boldsymbol{\theta}$, updating each component conditional on the latest values of all other components. The first b draws are discarded to ensure that only the draws taken after convergence are considered for the posterior distribution. For latent variable models, the Gibbs sampler is extended to include a data augmentation step, which treats the latent variables η as hypothetical missing data and samples these from their conditional distributions (Song & Lee, 2012). That is, observations from the joint posterior $p(\boldsymbol{\theta}, \eta \mid \boldsymbol{y})$ are simulated by alternately drawing from the conditional distributions $p(\boldsymbol{\theta} \mid \eta, \boldsymbol{y})$ and $p(\eta \mid \boldsymbol{\theta}, \boldsymbol{y})$ (Song & Lee, 2012). As the grouping of the components influences convergence, highly correlated elements of $\boldsymbol{\theta}$ should be updated simultaneously, i.e., in the same block (Jackman, 2009). For detailed information on the updating steps of the Gibbs sampler as implemented in Mplus see Asparouhov and Muthén (2010a). Gibbs sampling in Mplus is based on using conjugate priors (Asparouhov & Muthén, 2010a), that is, normally distributed priors for intercept, slope and loading parameters, inverse Wishart priors for variance-covariance matrices or an inverse gamma prior in case the updating block consists of only one parameter (see Section 7.3 for more details).

6.3 Convergence diagnostics and model fit

Model estimation via MCMC methods needs careful convergence diagnostics to ensure that the sampler has converged to the target distribution, i.e., the posterior distribution under consideration. This renders the inspection of convergence by use of convergence statistics and / or visual diagnostic tools such as trace plots indispensable whenever applying Bayesian MCMC estimation. The aim of convergence diagnostics is thus to check a) the representativeness of the sampled observations for the target distribution, i.e., they should be independent from the starting value and the Markov chain should not get stuck in subregions of the parameter space, b) that the Markov chain is long enough (sample size) to guarantee stable and reliable estimates and reduce Monte Carlo error to a minimum, and c) the efficiency of the MCMC simulation (Kruschke, 2015).

There are several ways to check convergence of the sampler to its equilibrium distribution. Visual inspection of the plotted simulated sequences of draws (traceplots) allow to check whether a) chains do not remain at the exact same value for a long sequence of draws, and b) several chains, started at different starting values, mix well together. The potential scale reduction factor (PSR, or Gelman-Rubin statistic, shrink factor; Gelman & Rubin, 1992) compares parameter variation within each chain to that across chains when running multiple independent Markov chains with different starting values. Convergence is assumed if the PSR falls below a certain value close to 1, i.e., if the PSR is less than $1+\varepsilon$ for all parameters in a model, where ε is chosen by the researcher and is most commonly set to a value between 0.05 and 0.1, depending on model complexity (Asparouhov & Muthén, 2010a).

Efficiency of the MCMC simulation is closely linked to the effective sample size of independent draws generated by sampling from the posterior distribution. The draws generated by a Markov chain are not independent, and the larger the dependency between successive draws the slower the chain explores the parameter space. Hence, with a high dependency between the draws in a Markov Chain, a larger number of iterations (draws) is needed to obtain a sufficient number of "effective" independent samples. Dependency between successive draws in the Markov chain can be estimated via the autocorrelation function of the chain. To reduce autocorrelation between the MCMC samples, only every k-th draw of the Markov chain can be used to generate the posterior distribution (thinning with factor k). While thinning does not increase efficiency of the simulation (and using the whole, unthinned chains provides more information and thereby accuracy), thinning can be useful when computer memory capacities might reach their limits in large models with very long chains (Link & Eaton, 2012).

Model fit in Mplus is assessed via posterior predictive p-values (PPP) using the difference between the observed and replicated Chi-Square values (Asparouhov & Muthén, 2010a). The posterior predictive distribution is the distribution of replicated data that can be drawn at every MCMC iteration step based on the posterior parameter estimates at that iteration. That is, draws from the posterior predictive distribution $p(\boldsymbol{y}^{rep} \mid \boldsymbol{y})$ can be generated based on the data, the model, and the parameter estimates at every iteration. Draws from the posterior predictive distribution are then used to evaluate model fit via the PPP value based on some test statistic T. The PPP is defined as the probability that the data generated under the model (the replicated data $\boldsymbol{y}^{rep}$) are more extreme than the observed data, as measured by test statistic T. That is, the PPP is estimated as the relative frequency with which $T(\boldsymbol{y}^{rep},\boldsymbol{\theta})$ is more extreme than $T(\boldsymbol{y},\boldsymbol{\theta})$ (Asparouhov & Muthén, 2010a; Gelman et al., 2014):

$$PPP = P(T(\boldsymbol{y}^{rep},\boldsymbol{\theta}) \geq T(\boldsymbol{y},\boldsymbol{\theta}) \mid \boldsymbol{y}) \approx \frac{1}{R}\sum_{r=1}^{R} I_{\{T(\boldsymbol{y}^{rep},\boldsymbol{\theta}_r)\geq T(\boldsymbol{y},\boldsymbol{\theta}_r)\mid \boldsymbol{y}\}} \tag{6.3.1}$$

In Mplus, every 10^{th} iteration after burn-in is used to compute the PPP (Asparouhov & Muthén, 2010a). In case of categorical data, the underlying continuous response variables $\boldsymbol{y}^*$ are used to compute the PPP. Values close to 0 and 1 indicate a discrepancy between model and data, while a PPP of 0.5 would indicate perfect model fit. However, there is no clear cut-off value to determine a well or ill fitting model based on the PPP. In practice, model fit with the PPP value is often evaluated based on a cut-off value of 5% (e.g., Asparouhov & Muthén, 2010b; Liang & Yang, 2014; van de Schoot et al., 2013), and, building on several simulation studies, B. Muthén and Asparouhov (2012)

conclude that using a PPP value of .10, .05 or 0.1 appears reasonable.
Several other convergence diagnostics (e.g., effective sample size) and model fit indices (e.g., Deviance information criterion) exist, are, however, not (yet) provided by Mplus (7.3) in case of categorical indicators and are therefore not discussed here. Detailed information on the use of the tools described above in the following simulation studies and application are given in Chapters 7 and 8.

Chapter 7

Monte Carlo Simulation Studies

7.1 Aims of the Monte Carlo Simulation Studies

To investigate the performance of the LS-Com GRM, LST-Com GRM and LGC-Com GRM as well as their applicability in different conditions, Monte Carlo simulation studies were conducted. The simulation studies were designed based on the results of the simulation studies on the LS-Com, LST-Com and LGC-Com models for continuous indicators (Koch et al., 2014). The objective of the simulation studies was to identify favorable and critical conditions for the application of the models as well as potential limits of their applicability. Differential effects of between- and within-level sample sizes on estimation accuracy were investigated to provide guidelines on the minimum required sample size in order to yield reliable and accurate parameter estimates. Furthermore, model complexity was considered by varying the number of constructs and measurement occasions. In practical applications of MTMM analyses, the degree of convergent validity between different methods may vary greatly depending on the research context. Possible effects of the degree of convergent validity on estimation accuracy were examined by including two different consistency conditions.

7.2 Previous results of simulation studies on sample size requirements

The simulation studies on the continuous-indicator counterparts of the models presented in this work have shown that at least 5 within-level observations and overall at least 5, better 10 observations per estimated parameter are required to obtain appropriate ML parameter estimates and to reduce the occurrence of improper solutions in the LS-Com, LST-Com and LGC-Com models (Koch, 2013; Koch et al., 2014, 2017). Previous recommendations for the (ML) estimation of continuous-indicator two-level SEMs comprised the use of a minimum of 100 between-level units (Hox & Maas, 2001; Julian, 2001; Maas & Hox, 2005; Meuleman & Billiet, 2009). Some simulation studies indicated a great influence of within-level sample sizes on parameter estimates (Koch, 2013; Koch et al., 2014; Yuan & Hayashi, 2005), while previous results with respect to continuous multilevel SEMs emphasized the importance of between-level sample sizes (Maas & Hox, 2005).
With respect to latent growth-curve models, Bishop et al. (2015) recommended to use sample sizes of at least 300 and at least four measurement time points to estimate indicator-specific growth models.
Using Bayesian estimation techniques, recommended sample sizes might vary from those needed for the estimation of continuous-indicator models estimated with maximum likelihood methods. Using Bayesian methods, estimation problems connected to improper solutions or inadmissible parameter

estimates can be avoided by assigning zero prior probability to these parameter spaces (Depaoli & Clifton, 2015; Hox et al., 2012). Furthermore, the applicability of complex models might be further enhanced by the possibility to incorporate informative prior information in the estimation process, especially in small samples (Depaoli & Clifton, 2015; Lee et al., 2010). Some studies have found Bayesian methods to outperform classical estimation methods (e.g., maximum likelihood) with regard to singlelevel (Lee & Song, 2004) and multilevel factor models with few clusters (Asparouhov & Muthén, 2010b; Hox et al., 2012). For instance, Bayesian as compared to ML estimation has been found to yield satisfactory results with only 20 instead of 50 to 100 clusters on the between level in continuous indicator multilevel models (Hox et al., 2012).

However, results on the estimation accuracy of Bayesian estimation with diffuse priors are mixed (Depaoli & Clifton, 2015; Holtmann et al., 2016; Hox et al., 2012). In a simulation study by Depaoli and Clifton (2015), Bayesian estimation with informative priors outperformed other estimation techniques with respect to the estimation of between-group parameters in models with dichotomous indicators, while diffuse priors led to biased between-group parameter estimates in small samples (< 100) and for low ICCs ($\leq .1$). For the less problematic within-group part, Bayesian estimation with diffuse priors required at least 100 level-2 observations for accurate estimates of loading parameters in dichotomous indicator models, while larger samples were needed for the accurate estimation of structural effects (Depaoli & Clifton, 2015).

The simulation study by Holtmann et al. (2016) found that larger sample sizes are needed for accurate estimation of the parameters in categorical-indicator multilevel SEMs when estimated with Bayesian methods using diffuse priors as compared to WLSMV estimation. Bayesian estimation only outperformed WLSMV estimation when applied with highly informative accurate priors. For a cross-sectional multi-construct multilevel model with interchangeable raters, a minimum of 150 between-level and 6 within-level observations were needed to obtain proper parameter estimates with Bayesian estimation with diffuse priors (Holtmann et al., 2016). Furthermore, larger sample sizes on both levels and more iterations until convergence are needed in order to obtain accurate Bayesian parameter estimates in categorical-indicator as compared to continuous-indicator SEMs (Depaoli & Clifton, 2015; Holtmann et al., 2016; Lee et al., 2010). Hence, the comparably more complex LS-Com, LST-Com and LGC-Com GRMs are expected to require larger sample sizes than their continuous-indicator counterparts, as well as larger sample sizes than the cross-sectional multilevel SEMs for interchangeable raters used in Holtmann et al. (2016).

7.3 Monte Carlo simulation designs

The simulation designs and specification of population parameters follow the simulation studies of the continuous indicator models (Koch, 2013) to ensure comparability. The simulated models include one or two constructs, measured by two methods, one structurally different method and one set of interchangeable methods. The models were simulated with two instead of three methods, as the simulation studies on the continuous indicator models have shown that the influence of including additional method factors on estimation accuracy is negligible (Koch, 2013; Koch et al., 2014). Furthermore, the LC-Com GRM was not simulated as the LS-Com GRM with strong measurement invariance and the baseline LC-Com GRM are mathematically equivalent (Koch, 2013) and hence similar results can be expected.

Three indicators were used per TMU. Six aspects were manipulated: (a) the number of constructs (1 or 2), (b) the degree of convergent validity / consistency (high vs. low), (c) the number of measurement occasions, (d) the number of level-1 units (i.e., raters per target), (e) the number of level-2 units (i.e., targets), and (f) the amount of prior information used in the estimation (diffuse vs. informative priors). Exact conditions are provided in the sections of the specific models.

Small sample size settings were chosen in order to investigate the model under minimal, realistic conditions. For instance, a review on applications of LST models in psychology and the social sciences revealed that these models were estimated with a median sample size of 249 observations in practice (Geiser & Lockhart, 2012). Given that the MTMM GRM models simulated in the present studies are likely to be more complicated, the inclusion of larger sample sizes seems appropriate. The population model parameters were chosen based on the variance decomposition they yielded in the respective model, that is, the degree of consistencies and method specificities. Population values for the different variance components in the two consistency conditions are provided in the sections of the respective model.
The simulations comprised 200 replications per condition. Data was generated with three response categories per item, using a probit link and the Theta parameterization in Mplus (Asparouhov & Muthén, 2007). Data sets were simulated using Mplus 7.3 (L. K. Muthén & Muthén, 1998-2012), models were estimated using Mplus 7.3 (L. K. Muthén & Muthén, 1998-2012), and results were analyzed using the software R 3.0.2 (R Development Core Team, 2013) as well as the R package "MplusAutomation" (Hallquist & Wiley, 2014). Posterior parameter estimates were obtained using Bayesian estimation methods with Mplus 7.3 (L. K. Muthén & Muthén, 1998-2012). Extensive pre-analyses were used to determine an appropriate number of iterations, burn-in samples and thinning. Trace-plots of MCMC samples, autocorrelation plots and PSR values were used to examine convergence and dependency of the MCMC draws. The exact numbers of iterations and thinning used are given in the subsection on the respective model.
In the diffuse prior condition, prior specifications were left to the Mplus default settings. This was done assuming that it is the approach adopted by most applied researchers. These prior settings correspond to

$$\lambda \sim N(0,5) \tag{7.1.1}$$
$$\kappa \sim N(0,10^{10}) \tag{7.1.2}$$
$$\mu \sim N(0,10^{10}) \tag{7.1.3}$$
$$\Phi_W \sim IW(\mathbb{I}_m, m+1) \tag{7.1.4}$$
$$\Phi_B \sim IW(\mathbb{I}_\nu, \nu+1) \tag{7.1.5}$$

where Φ_W represents the within-level variance covariance matrix, Φ_B represents the between-level variance covariance matrix, N denotes the density of the normal distribution, IW the Inverse Wishart distribution, and $\mathbb{I}_m$ and $\mathbb{I}_\nu$ are Identity matrices of size $m \times m$ and $\nu \times \nu$, respectively, where m and ν correspond to the size of the covariance matrices. Note that the size of the covariance matrices Φ_W and Φ_B depend on the model and the number of measurement occasions and constructs included in the model. The prior $N(0,10^{10})$ approximates a constant uniform prior on the interval $(-\infty,\infty)$.
In the informative prior condition, informative priors were set on loadings, thresholds and means. They were given normal priors, with the prior mean corresponding to the respective parameter's value in the data generation and a prior variance of 0.1. A prior variance of 0.1 (prior *SD* = 0.32) implies a prior distribution where the central 99% of the values lie in the interval [mean - 0.62; mean + 0.62].
Note that the simulation studies addressed the frequentist properties of the Bayesian parameter estimates. In this context, the term *population parameters* refers to the parameter values used for data simulation. The mean of the posterior distribution was used as a point estimate and posterior quantiles were used for providing a 95% credibility interval for parameter estimates.
There is an ongoing debate on whether to exclude replications that show convergence problems or suffer from the presence of improper parameter estimates (Boomsma, 2013; Chen, Bollen, Paxton, Curran, & Kirby, 2001). Improper solutions are in general less an issue in Bayesian estimation than in classical estimation techniques, as they can be prevented by assigning zero prior probabilities to the respective parameter values. Hence, improper solutions such as negative variances or correlations

greater than 1 often encountered in these kind of complex models with ML estimation (see, e.g., Koch, 2013, for rates of improper solutions occurring in the respective continuos indicator models), cannot occur in the present simulation studies, as prior settings of the Bayesian estimation exclude these kind of parameter values. Convergence problems, on the other hand, might pose a big problem in certain conditions. PSR values alone are not considered a valid criterion for excluding replications from further analyses, as replications with high PSR values for one parameter might show better convergence for many other parameters than a replication that has a PSR value just below the chosen cut-off. As it is difficult to decide on convergence by objective criteria or a clear cut-off value in a simulation study using Bayesian estimation, excluding replications that are considered non-converged by these criteria might render interpretability of the results more difficult. Furthermore, applied researchers might not discard their results based on PSR values, especially in cases where convergence problems might be restricted to single parameters. Hence, excluding these replications might threaten external validity. Therefore, we decided to include all requested replications in the analysis of our Monte Carlo simulations of Bayesian estimators. For conditions with large rates of replications showing convergence problems, results should be interpreted with caution.

7.4 Evaluation criteria

To evaluate the performance of the models the following criteria were used: (a) the PSR value after the requested number of iterations (as an indicator of convergence), (b) model fit as indicated by PPP values using the difference between the observed and replicated Chi-square values, (c) the amount of parameter estimation and standard error bias, and (d) the coverage of the parameter's data generating value by 95% credibility intervals (CIs).
Bias for parameter p of parameter class c was calculated as

$$Bias_{pc} = \frac{1}{n_{rep}} \sum_{e=1}^{n_{rep}} \widehat{\theta}_{pce} - \theta_{pc} \tag{7.2.1}$$

where $\widehat{\theta}_{pce}$ is the parameter estimate of replication e of parameter p belonging to parameter type c, θ_{pc} is the true, data generating value of the respective parameter and n_{rep} is the number of replications (Bandalos, 2006). The relative parameter estimation bias (peb) was calculated for each parameter and then averaged over parameters of the same parameter type (e.g. state loadings) in the following way:

$$peb(c) = \frac{1}{n_c} \sum_{c=1}^{n_c} \left| \frac{\frac{1}{n_{rep}} \sum_{e=1}^{n_{rep}} \widehat{\theta}_{pce} - \theta_{pc}}{\theta_{pc}} \right| \tag{7.2.2}$$

with n_c as the number of parameters in parameter class c. In line with L. K. Muthén and Muthén (2002) and Koch et al. (2014), peb values falling below a cut-off value of 0.10 (10%) were considered acceptable. Peb values falling between 0.10 and 0.30 (10% and 30% deviation from the population value) were considered as medium and peb values > 0.30 as large bias.
Standard error bias per parameter was calculated by

$$(SE - SD)(pc) = \frac{1}{n_{rep}} \sum_{e=1}^{n_{rep}} \widehat{sd}(\widehat{\theta}_{pc})_e - sd(\widehat{\theta}_{pc}) \tag{7.2.3}$$

with $\widehat{sd}(\widehat{\theta}_{pc})_e$ being the posterior standard deviation (SD) of parameter $\widehat{\theta}_{pc}$ at replication e, and $sd(\widehat{\theta}_{pc})$ the empirical SD of the parameter estimates over all replications (Bandalos, 2006). The relative standard error bias (seb) was calculated for each parameter and then averaged over parameters of the same parameter type (e.g. state loadings) in the following way:

$$seb(c) = \frac{1}{n_c}\sum_{c=1}^{n_c}\left|\frac{\frac{1}{n_{rep}}\sum_{e=1}^{n_{rep}}\widehat{sd}(\widehat{\theta}_{pc})_e - sd(\widehat{\theta}_{pc})}{sd(\widehat{\theta}_{pc})}\right| \tag{7.2.4}$$

Again, seb values falling below a cut-off value of 0.10 (10%) were considered acceptable. Note that in the context of the simulation study, posterior standard deviations ($\widehat{sd}(\widehat{\theta}_{pc})_e$) were used as an analogue of a Bayesian standard error (SE).
Additionally, empirical SDs will be considered, which, in combination with the bias values, can be used as an indicator of the mean squared error (MSE):

$$MSE_{pc} = sd(\widehat{\theta}_{pc})^2 + Bias_{pc}^2 \tag{7.2.5}$$

95% coverage is the proportion of replications for which the 95% CI contains the true parameter value. Coverage between 92% and 98% was considered acceptable.

7.5 Monte Carlo simulation LS-Com GRM

7.5.1 Simulation design

The model used in the simulation study is the LS-Com GRM with common state variables, depicted in Figure 7.1. Five aspects were manipulated in the simulation[1]: (a) the number of constructs (1 or 2), (b) the degree of convergent validity / consistency (high vs. low), (c) the number of measurement occasions (2, 3, and 4), (d) the number of level-1 units (i.e., raters per target: nL1 = 2, 5, 10, and 20), (e) the number of level-2 units (i.e., targets: nL2 = 250, 500, and 750), and (f) the amount of prior information used in the estimation (diffuse vs. informative priors), resulting in a 2x2x3x4x3x2 design. In total, the simulation study was comprised of 288 conditions and was simulated with 200 replications per condition. The population model parameters were chosen based on the variance decomposition in the LS-Com GRM, that is, the degree of consistency and method specificity. Population values for the different variance components in the two consistency conditions are given in Table 7.1. Population values for all parameters in the LS-Com GRM data generation can be found in Table A 1 in Appendix A.1.
All models were specified and data generated with common latent state variables, assuming strict factorial invariance (Meredith, 1993; Meredith & Teresi, 2006). For identification reasons the first loading per factor was set to 1 and the latent state means of the first measurement occasion were set to 0. Extensive pre-analyses were used to determine an appropriate number of iterations, burn-in samples and thinning. Trace-plots of MCMC samples and PSR values were used to examine convergence. Based on these analyses, Bayesian estimation of the LS-Com GRM was conducted running 2 MCMC chains, with 60,000 sample iterations per chain, using a thinning factor of 3. That is, 30,000 burn-in iterations and 30,000 iterations after burn-in were run per chain, using only every third iteration as a sample for the posterior distribution.
Estimation times per replication lay between 3.5 minutes for the smallest model (one construct, two measurement occasions, nL2 = 250, nL1 = 2) and 6 hours for the largest model (two constructs, four measurement occasions, nL2 = 750, nL1 = 20). Simulating on 15 computers, this resulted in an approximate estimation time of 5 months for the simulation study on the LS-Com GRM.

[1]The simulation study on the mono-construct LS-Com GRM has been pre-published in Holtmann et al. (2017)

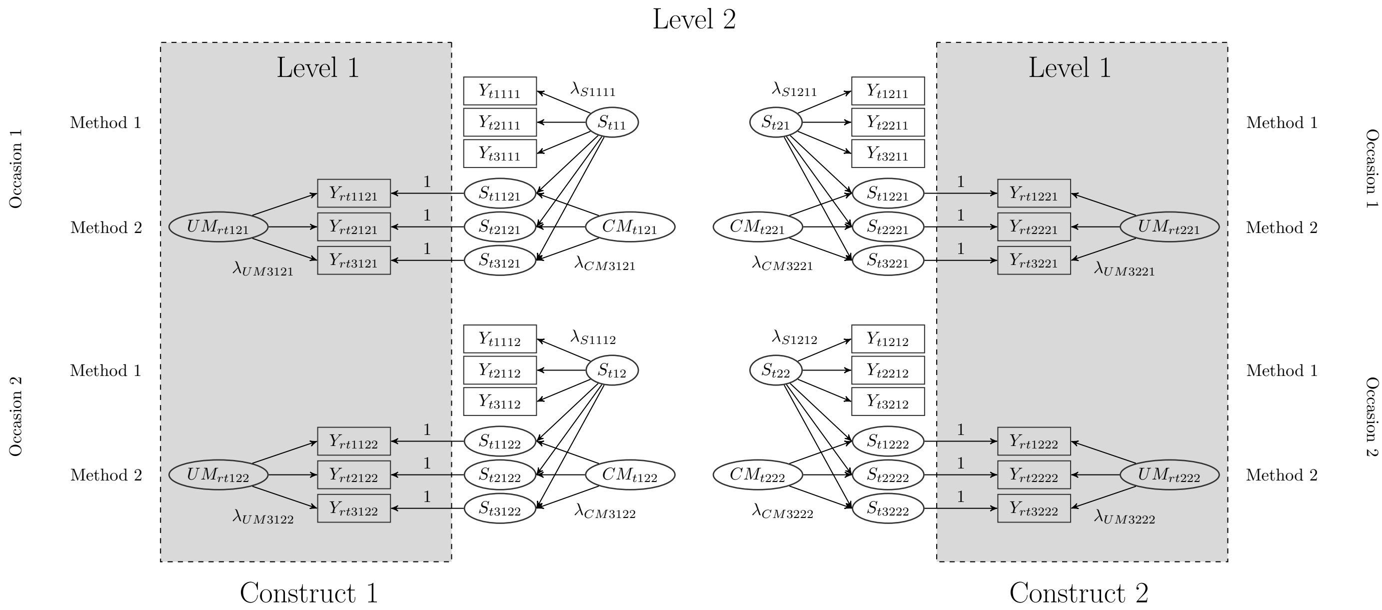

Figure 7.1: Path diagram of the Latent-State-Com GRM with common state variables S_{tjl} and two constructs measured by two methods on two measurement occasions, as simulated in the LS-Com GRM simulation study. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are probabilistically linked to the latent variables by a probit link. For the sake of clarity, correlations between latent variables were omitted and loading parameters are only shown for exemplary indicators. Correlations that were set to zero are correlations between latent state and latent common methods variables, and correlations between any level-1 and any level-2 latent variable. *CM*: common method variable; *M*: method variable; *S*: latent state variable; *UM*: unique method variable; Y_{rtijkl}: observed variable for the rating of rater *r* for target *t* of the *i*-th item of trait *j* and method *k* on occasion *l*.

Table 7.1: Population parameters of the consistency and method specificity coefficients and of the latent factor correlations in the LS-Com GRM simulation study

Consistency and method specificity coefficients					
		Low consistency		High consistency	
Coefficient		Mean	*SD*	Mean	*SD*
Consistency		0.375	0.016	0.750	0.006
UM specificity		0.312	0.018	0.124	0.023
CM specificity		0.313	0.034	0.126	0.029
Latent correlations					
			measurement occasion		
Factor	Construct	$l = l'$	$l - l' = 1$	$l - l' = 2$	$l - l' = 3$
State	same		0.6	0.5	0.4
	different	0.5	0.3	0.2	0.1
UM and *CM*	same		0.6	0.5	0.4
	different	0.3	0.1	0.1	0.1

Note. Displayed are the mean consistency and method specificity coefficients over the different non-reference method items in the high and low consistency conditions, as well as their variation in standard deviations (*SD*). *CM*: Common method; *l*: occasion of measurement; *UM*: Unique method.

7.5.2 Results

Convergence

Figure 7.2 displays the distributions of the PSR values at the last iteration over the different replications per condition. Despite the high number of burn-in and sampling iterations used, PSR values higher than 1.1 still occurred, primarily in the high consistency condition with only 2 (or 5) observations on the within-level, when estimated with diffuse priors. The parameters associated with these high PSR values were primarily loadings of the state and common method factors on the interchangeable peer report indicators. Convergence as judged by PSR and trace plots was good in all other conditions.

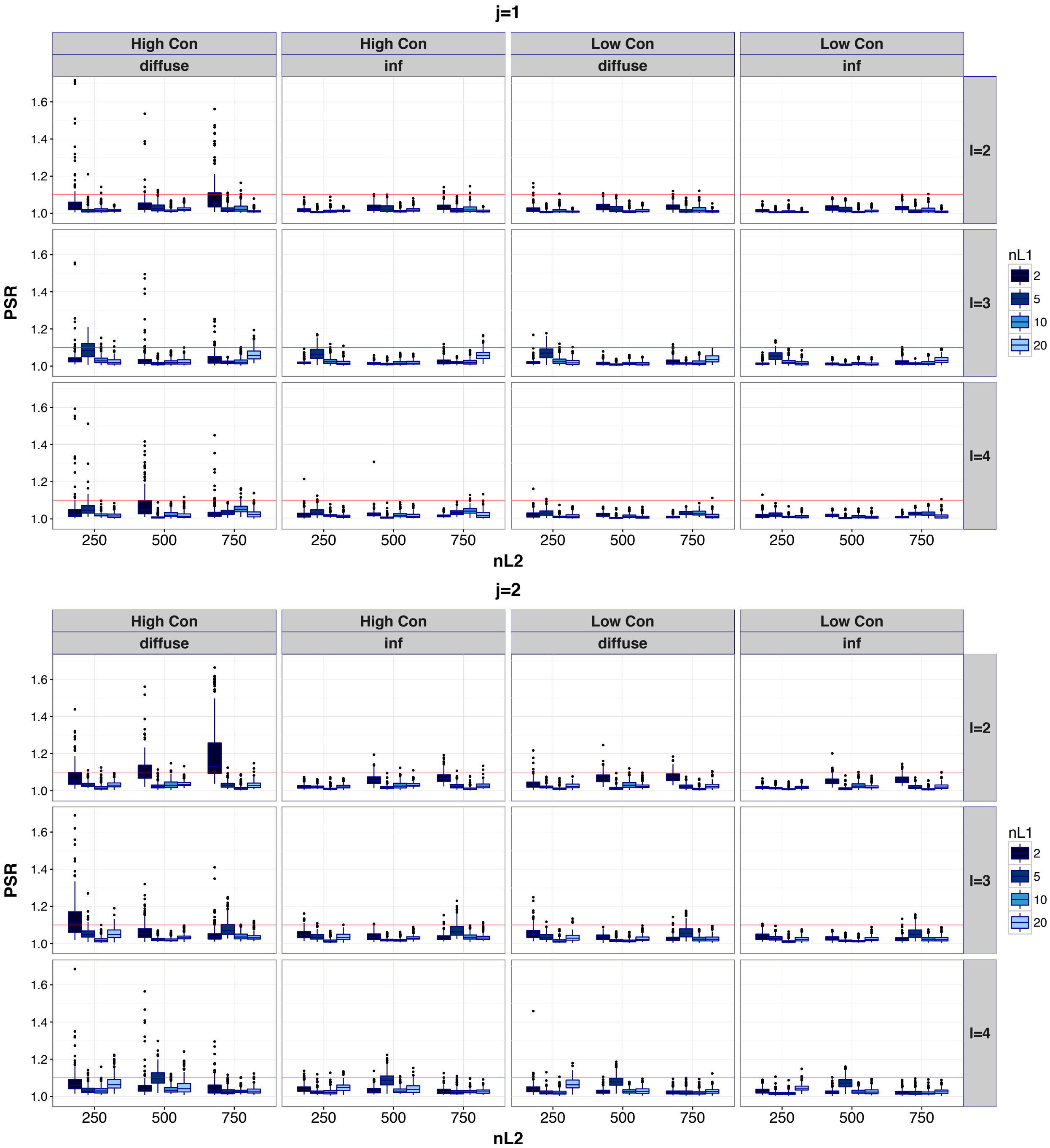

Figure 7.2: Boxplots of Potential scale reduction (PSR) values at the last iteration over all 200 replications per condition in the LS-Com GRM simulation study, for the mono-construct condition ($j = 1$) and the multi-construct condition ($j = 2$). Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations; PSR: Potential scale reduction.

Model fit

In the mono-construct models ($j = 1$) PPP values all fell below the value of .9, and PPP values smaller than .1 were observed in 2.8% of the conditions. These were high consistency conditions with large numbers of observations, each with 0.5% of the replications having a PPP below .1. PPP values between .1 and .2 occurred in 27.0% of the conditions, mainly in the conditions with 4 measurement occasions, high consistency and large sample sizes. A maximum of 1% of the replications per condition in the mono-construct models had a PPP between .1 and .2. In the mono-construct models, a minimum of 88.5% up to a maximum of 100% of the PPP values per condition fell into the range between .3 and .7.

In the multi-construct models ($j = 2$) PPP values all fell into the range between .1 and .9, that is, no PPP value smaller than .1 was observed in any replication of any condition. PPP values between .1 and .2 occurred in 24.3% of the conditions, mainly in the conditions with 2 measurement occasions, high consistency and large sample sizes. A maximum of 2% of the replications per condition in the multi-construct models had a PPP $< .2$. Over all replications and conditions of the multi-construct models, a minimum of 87.5% up to a maximum of 100% of the PPP values per condition fell into the range between .3 and .7.

Recall that a PPP $< .1$ or $< .05$ would indicate poor model fit, according to the recommendations given by B. Muthén and Asparouhov (2012).

Parameter estimation bias and coverage.

Mean peb values averaged over all model parameters per condition are displayed in Figure 7.3. It is apparent that on average bias was higher in the high consistency condition as compared to the low consistency condition, with peb values for high consistency conditions with few observations exceeding the cut-off of 10%. On average, parameters in the low consistency conditions are estimated with acceptable accuracy in both the mono- and the multi-construct conditions. Because of considerable variability in bias across parameter classes, accuracy of parameter estimation is described separately for each parameter class in the following.

Biases per parameter, mean coverage, peb and mean MSE values for the parameters in the LS-Com GRM are displayed in Figures A 1 - A 6 for the mono-construct condition in Appendix A.2, and Figures A 12 - A 17 for the multi-construct condition in Appendix A.3.

Mean absolute bias for the state loadings was smaller than 0.105 in all conditions, that is, deviations were less than 12.4% from the population parameters. State loading coverage values showed most problems for the multi-construct model in the low consistency condition with 4 measurement occasions. Unique and common method loadings were estimated with high accuracy in the low consistency condition, with biases less than 7.9% and 11.6% of their population values, respectively. In the high consistency condition, however, common method loadings exhibited peb values ranging from 0.058 up to 0.599, biases being highest in the conditions with few within-level observations and diffuse priors. Common method loading coverage did not reach 80% in most cases of the high consistency conditions, with the lowest coverage values in conditions with 4 measurement occasions. Unique method loading biases in the high consistency condition revealed a similar pattern, with somewhat lower peb values (up to 0.287 for $j = 2$ and 0.367 for $j = 1$) and higher coverage ($> 81.6\%$ for $j = 2$ and $> 66.5\%$ for $j = 1$).

For the latent state variances, peb values lay between .001 and .153, with peb values $> .10$ occurring almost solely in conditions with few between-level units (nL2 = 250). Coverage ($81.3\% <$ coverage $< 98.3\%$) was lowest for the multi-construct condition with 4 measurement occasions. While for the state variances bias increases and coverage decreases with an increasing number of measurement occasions, this effect is not observed for unique or common method variances. Variance estimates of the method factors are estimated accurately in the low consistency condition ($0.003 <$ peb < 0.385; and

92.8% < coverage < 98.1%), with peb values > .10 occurring only in the nL1 = 2 or nL1 = 5 with nL2 = 250 conditions. In the high consistency condition, however, method factor variance estimates showed higher biases (0.089 < peb < 1.999; and 16.5% < coverage < 91.0%).
Bias in the estimation of the covariance parameters is considerably small and coverage values lay in the desired range with few exceptions (see Figures A 5, A 6, A 16, and A 17).
Threshold parameters and latent state means were estimated accurately, with mean peb values < 3.4% and < 10.5%, respectively.

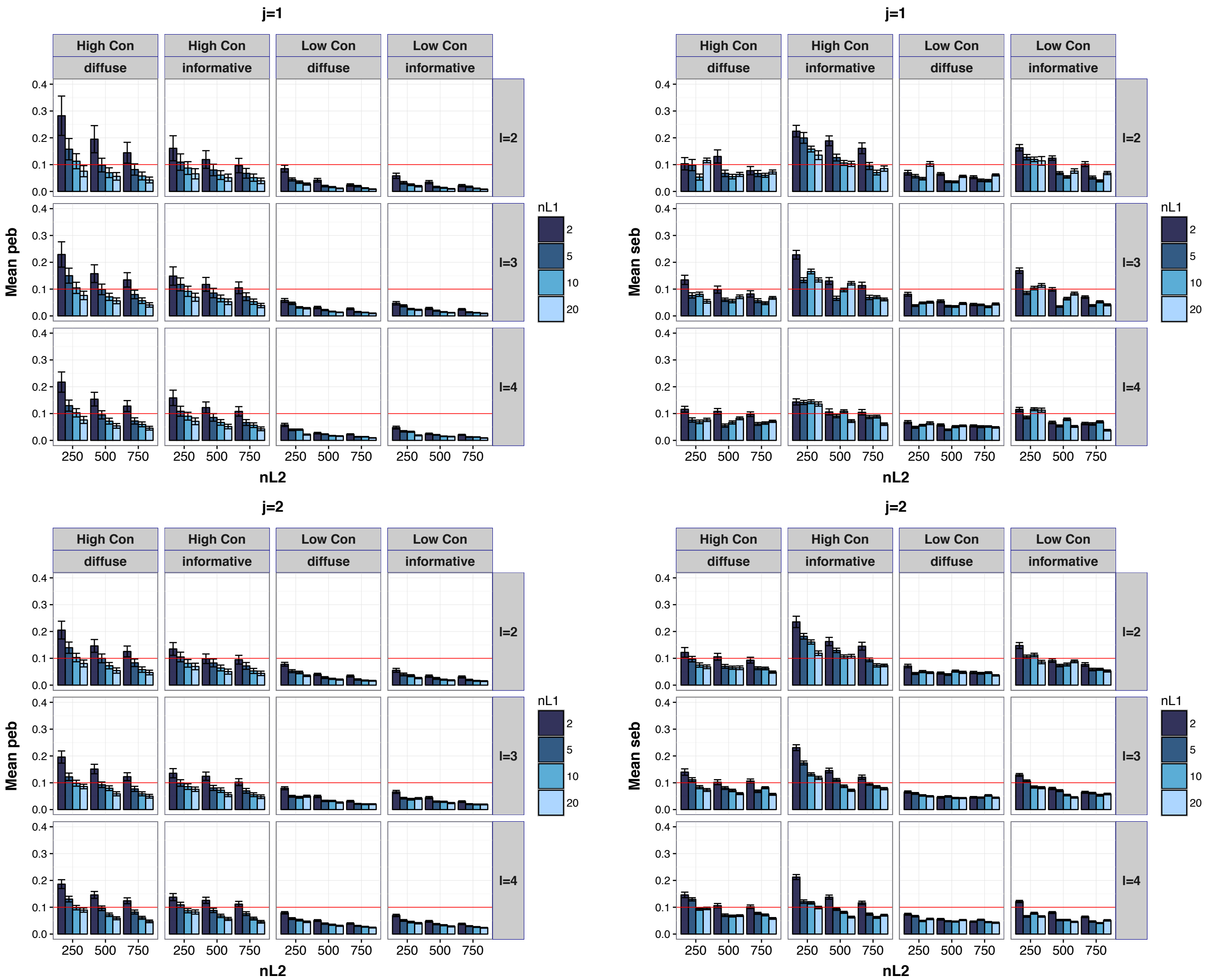

Figure 7.3: Mean parameter estimation bias (peb; left hand side) and standard error bias (seb; right hand side) values averaged over all parameters per condition of the mono-construct model ($j = 1$; upper panel) and the multi-construct model ($j = 2$; lower panel) in the LS-Com GRM simulation study. Error bars represent standard errors. Diffuse: diffuse prior condition; High Con: high consistency condition; informative: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

Standard error bias.

Mean seb values averaged over all model parameters per condition are displayed in Figure 7.3. It is apparent that on average standard error bias was higher in the informative prior conditions as compared to the diffuse prior conditions, with seb values for the informative prior conditions with few

observations mostly exceeding the cut-off of 10%. On average, standard error bias is acceptable in the diffuse prior conditions. In the following, accuracy in the estimation of posterior *SD*s is described separately for each parameter class.
Differences between empirical and posterior *SD*s per parameter as well as mean empirical *SD*s for the parameters in the LS-Com GRM are displayed in Figures A 7 - A 9 for the mono-construct condition in Appendix A.2, and Figures A 18 - A 20 for the multi-construct condition in Appendix A.3.
As theoretically expected, empirical *SD*s decrease with increasing number of observations on both the within- and between-level for all parameter types and conditions. Informative priors lead to smaller empirical *SD*s primarily for the loading parameters in conditions with few within-level observations. Dot plots of standard error bias for the loading parameters show that bias is higher in the informative as compared to the diffuse prior conditions, with posterior *SD*s mainly overestimating empirical *SD*s when informative priors are used. A similar pattern can be observed for the standard error bias of the state variances, where the tendency of overestimation subsides with an increasing number of measurement occasions. In contrast, posterior *SD*s of the method variance parameters overestimate empirical *SD*s with both diffuse and informative priors in the high consistency condition, especially in combination with few observations on both the within- and between-level. Note that no informative priors were set on variance parameters. Bias is slightly smaller in the low consistency conditions. Patterns of deviations of average posterior *SD*s from empirical *SD*s for the covariance parameters resemble those of the variance parameters, while deviations are smaller for the covariance estimates as compared to the variance estimates in absolute value.

7.5.3 Summary and conclusion

The results of the simulation study show that the parameters of the LS-Com GRM can be accurately estimated with Bayesian estimation methods for a level of convergent validity that is typically found in practice (consistency around .375). To further reduce potential bias in parameter estimates, it seems recommendable to sample more than 2 and more than 250 observations on the within- and between-level, respectively. High PSR values (> 1.1) in some of the conditions with only two observations on the within-level provide additional evidence that two raters per target might be insufficient. Due to the large number of iterations chosen, these results may be indicative of insufficient empirical information on the within-level with nL1 = 2.
The results corroborate the recommendation for the continuous LS-Com model to sample more than two raters per target (Koch et al., 2014) and of categorical MTMM models to sample more than 4 raters per target for Bayesian estimation with diffuse priors (Holtmann et al., 2016). In contrast to the continuous indicator LS-Com model, more observations (approximately 250 instead of 100) are required on the between-level. This result supports previous findings on singlelevel CFA-MTMM models with ordinal variables reporting a need of at least 250 observations when two to four indicators are used per TMU (Nussbeck et al., 2006).
The simulation results show that the number of level-1 units has a substantial impact on estimation accuracy of the within-level parameters and, similar to a result by Depaoli and Clifton (2015), the number of level-1 units and the degree of convergent validity (or ICCs) had the largest impact on convergence rates. These results correspond to the results of the simulation study by Koch et al. (2014) on the continuous indicator LS-Com model as well as the results of Bayesian estimation of categorical-indicator multilevel SEMs in Holtmann et al. (2016).
The estimation accuracy of state loadings and state variances decreased with increasing the number of measurement occasions. Estimation accuracy is acceptable for models with 2 or 3 measurement occasions when using the recommended sample sizes.
In case of high convergent validity, estimation of the method loading and method variance parameters of the LS-Com GRM appears to be more problematic. Estimation accuracy can, however, be im-

proved by increasing the number of observations on both levels. This corresponds to the findings for the continuous indicator LS-Com model (Koch et al., 2014), where parameter estimates were more accurate in cases of low as compared to high convergent validity. Low convergent validity is often encountered in practice so that generally unbiased results are to be expected with these moderate sample sizes.

Furthermore, the results show that if only few observations are available, (accurate) informative priors can be effectively used to decrease bias in the LS-Com GRM. This is true not only for the parameters that were given informative priors (e.g., the loadings), but also to smaller degrees for the remaining parameters (e.g., the variances). The chosen degree of prior information (a prior variance of 0.1) did, however, not affect the results in cases where sufficient information in terms of observations was provided. This is in line with theoretical considerations as well as previous results reporting a decrease of the prior's influence with increasing sample size (Asparouhov & Muthén, 2010b; Lee et al., 2010). Standard error bias was highest for the method loading parameters on both levels, did, however, decrease with increasing sample sizes on both levels. Interestingly, standard error bias increased with the use of informative priors, with average posterior *SD*s mostly overestimating empirical *SD*s. See section 7.8 for a discussion on standard error bias and results on model fit as indicated by PPP-values. Overall, the results of the simulation study are encouraging and indicate that the LS-Com GRM can be applied in a wide range of situations with relatively few observations.

7.6 Monte Carlo simulation LST-Com GRM

7.6.1 Simulation design

The model used in the simulation study is the LST-Com GRM with indicator-specific trait as well as method trait variables but common state residual variables, depicted in Figure 7.4. Five aspects were manipulated in the simulation: (a) the number of constructs (1 or 2), (b) the degree of convergent validity (high vs. low), (c) the number of measurement occasions (2, 3, and 4), (d) the number of level-1 units (i.e., raters per target: nL1 = 2, 5, 10, and 20), (e) the number of level-2 units (i.e., targets: nL2 = 250, 500, and 750), and (f) the amount of prior information used in the estimation (diffuse vs. informative priors), resulting in a 2x2x3x4x3x2 design. In total, the simulation study was comprised of 288 conditions and was simulated with 200 replications per condition. The population model parameters were chosen following the simulation study of the continuous indicator LST-Com model, based on the variance decomposition in the LST-Com GRM, that is, the degree of occasion-specificity, trait- and occasion-specific consistencies and method specificities. Population values for the different variance components of the high and the low consistency conditions are given in Table 7.2. Population values for all parameters in the LST-Com GRM data generation can be found in Table B 1 in Appendix B.1.

All models were specified assuming strong measurement invariance for all items over measurement occasions (Meredith, 1993; Meredith & Teresi, 2006). For identification reasons the first loading per factor was set to 1 and the latent trait means were set to 0. Note that, as a consequence, there are no estimates of unique or common method trait loadings, as these were set to 1 for all measurement occasions due to measurement invariances assumptions. Additionally, there are no covariances to be estimated for state residual, unique method state residual or common method state residual variables in the mono-construct model ($j = 1$), as covariances for these variables would only be permissible between different constructs (state residual correlations over time were set to zero for parsimony and identifiability reasons).

Extensive pre-analyses were used to determine an appropriate number of iterations, burn-in samples and thinning, using trace-plots of MCMC samples and PSR values. Based on these analyses, Bayesian

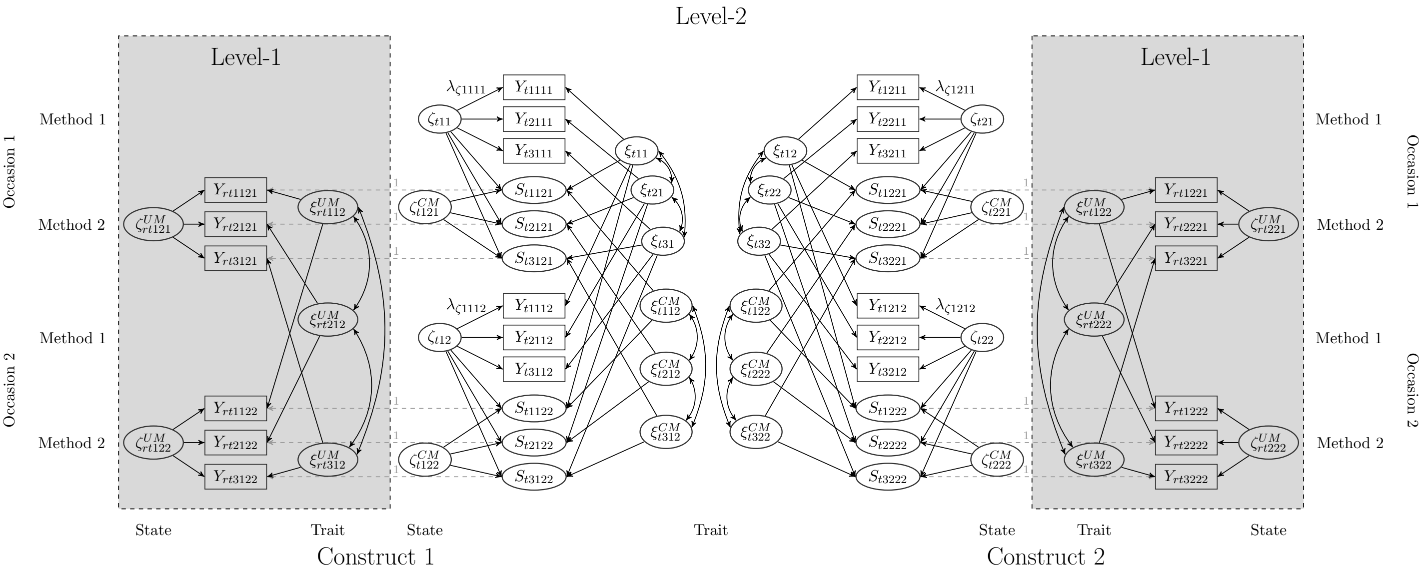

Figure 7.4: Path diagram of the Latent-State-Trait-Com graded response model with indicator-specific latent trait variables ξ_{tij} and ξ^{M}_{tijk} and common latent state residual variables ζ_{tjl}, for one structurally different method and one set of interchangeable methods on two measurement occasions for two constructs, as simulated in the LST-Com GRM simulation study. Method 1 is selected as reference method. For convenience, the constant indicator $k = 1$ has been dropped from the latent trait variables (ξ_{tij}=ξ_{tij1}) and the latent state residual variables (ζ_{tjl}=ζ_{tj1l}). For the sake of clarity, correlations between latent variables of different constructs are omitted and loading parameters are only shown for exemplary indicators. All correlations that were not constrained to zero between latent variables of the same construct are depicted by double-headed arrows. Other correlations that were not constrained to zero are: correlations between the latent traits ξ; between the variables ξ^{CM}; between the variables ξ^{UM}; between the latent state residual (method) variable of the first construct with the same variable of the second construct of the same measurement occasion. *CM*: common method; *M*: method; *S*: state variable; *UM*: unique method; ξ: latent trait variable; Y_{rtijkl}: observed variable for the rating of rater *r* for target *t* of the *i*-th item of trait *j* and method *k* on measurement occasion *l*; ζ: latent state residual variable.

Table 7.2: Population parameters of consistency, occasion specificity and method specificity coefficients and of the latent correlations in the LST-Com GRM simulation study

Variance components		
Coefficient	Low consistency	High consistency
Trait specificity coefficients		
Reference method	.5765	.5765
Non-reference method	.5000	.5000
Trait *CM* specificity	.3025	.1503
Trait *UM* specificity	.3951	.1503
Trait consistency	.3025	.6994
Occasion-specificity coefficients		
Reference method	.4235	.4235
Non-reference method	.5000	.5000
Occasion-specific *CM* specificity	.3025	.1503
Occasion-specific *UM* specificity	.3951	.1503
Occasion-specific consistency	.3025	.6994
Latent correlations		
	Construct	
Factor	same	different
Traits	.8	.4
State residuals	-	.2
CMT	.6	.3
UMT	.6 / .3	.369 / .185
CMS	-	.150
UMS	-	.150 / .092

Note. Displayed are the consistency, occasion specificity and method specificity coefficients (as defined in Table 4.1) over the different items as well as factor correlations used in the LST-Com GRM simulation study for the high and low consistency conditions. *CM*: Common method; *CMS*: Common method state residual; *UMS*: Unique method state residual; *CMT*: Common method trait; *UM*: Unique method; *UMT*: Unique method trait.

estimation of the LST-Com GRM was conducted running 2 MCMC chains and using a thinning factor of 5. The iterations per chain were varied by condition, based on convergence behavior of the chains in the respective condition (increasing iterations for lower sample sizes and higher model complexity) and estimation times (not increasing iterations for conditions where not necessary). Iterations were fixed to 100,000 in the informative prior conditions. That is, 50,000 burn-in iterations and 50,000 iterations after burn-in were run per chain, using only every fifth iteration as a sample for the posterior distribution. Iterations used in the diffuse prior conditions are given in Table 7.3.

Estimation times per replication lay between 7.1 minutes for the smallest model (informative priors, one construct, two measurement occasions, nL2 = 250, nL1 = 2) and 14.3 hours for the largest model (diffuse priors, two constructs, four measurement occasions, nL2 = 750, nL1 = 20). Simulating on 15 computers, this resulted in an approximate estimation time of 7 months for the simulation study on the LST-Com GRM.

Table 7.3: Number of iterations used in the diffuse prior conditions for the LST-Com GRM simulation study.

		One construct			**Two constructs**		
nL2	nL1	$l = 2$	$l = 3$	$l = 4$	$l = 2$	$l = 3$	$l = 4$
250	2	300	300	300	200	200	200
	5	300	300	300	200	200	150
	10	300	300	300	200	150	100
	20	300	300	300	200	100	100
500	2	300	300	300	200	200	200
	5	300	300	300	200	150	150
	10	300	200	200	150	100	100
	20	300	200	200	150	100	100
750	2	300	300	300	200	200	200
	5	300	200	200	200	150	150
	10	200	200	200	150	100	100
	20	200	200	200	150	100	100

Note. Iteration numbers are given in thousands. For an entry of x iterations, x/2 *1000 burn-in iterations and x/2 *1000 iterations after burn-in were run per chain, using only every fifth iteration as a sample for the posterior distribution (i.e., with 2 chains, posterior distribution samples were of size x/5*1000). l: number of measurement occasions; nL1: number of level-1 observations; nL2: number of level-2 observations.

7.6.2 Results

Convergence

Figure 7.5 displays the distributions of the PSR values at the last iteration over the different replications per condition. In the mono-construct condition (j = 1), the estimation exhibited immense convergence problems under high consistency with diffuse priors, with only very few exceptions in conditions with large sample sizes on both levels. Convergence problems could be reduced by the use of informative priors, did, however, not entirely disappear in the high consistency condition. In the low consistency condition of the mono-construct models, convergence problems were restricted to conditions with only 2 observations on the within-level or with 5 observations on the within-level in combination with few observations on the between-level or few measurement occasions. Parameters that showed most convergence problems in these conditions were state residual and trait loadings on the interchangeable informant report items, threshold parameters of the interchangeable informant report items and in the high consistency condition also the variances of the unique method state residuals. Results concerning the respective conditions should be interpreted with considerable caution. In the multi-construct condition convergence was mainly an issue in the high consistency condition

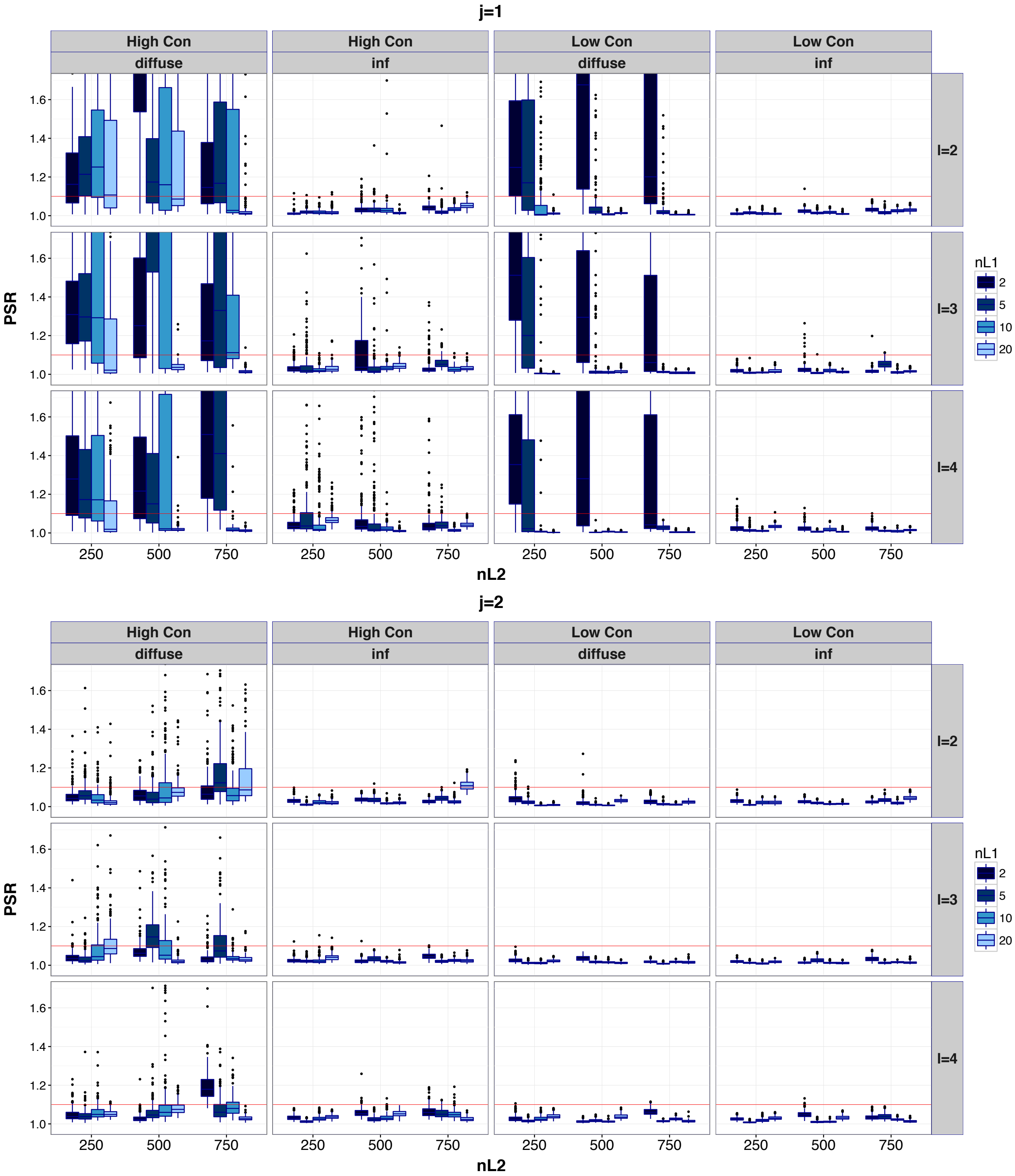

Figure 7.5: Boxplots of Potential scale reduction (PSR) values at the last iteration over all 200 replications per condition in the LST-Com GRM simulation study, for the mono-construct condition ($j = 1$) and the multi-construct condition ($j = 2$). Note that y-axes are only displayed to a maximum value of 1.75 to enhance readability of the plots in the relevant range. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations; PSR: Potential scale reduction.

with diffuse priors, however to much lesser degrees than in the mono-construct condition. Parameters with highest PSR values in these conditions were state residual and trait loadings on the interchangeable informant report items as well as the variances of the unique method state residuals. All other conditions of the multi-construct models (low consistency, informative priors) showed good convergence rates as judged by PSR and trace plots.

Model fit

PPP values were $< .9$ in all replications over all conditions. PPP values smaller than .1 were observed in 4.86% of the mono-construct and 2.80% of the multi-construct conditions. These were primarily high-consistency conditions with large within-level sample sizes. 0.5% to 1% of the replications in the respective conditions had PPP values smaller than .1. Recall that a PPP $< .1$ or $< .05$ would indicate poor model fit, according to the recommendations given by B. Muthén and Asparouhov (2012). PPP values between .1 and .2 occurred in 66.0% of the mono-construct conditions, with 0.5% to 7.5% of the replications per condition showing values in this range and higher percentages in conditions with larger within-level sample sizes. In the multi-construct models, PPP values between .1 and .2 occurred in 45.5% of the conditions, mostly in high consistency conditions (with 0.5% - 5.0% of the replications per condition showing values in this range). Between 0.5% and 31.0% and between 0.5% and 21.0% of the PPP values per condition lay between .2 and .3 in the mono-construct and multi-construct models, respectively. In the mono-construct models, a minimum of 60.5% up to a maximum of 99.5%, and in the multi-construct models, a minimum of 76.0% up to a maximum of 100% of the PPP values per condition fell into the range between .3 and .7.

Parameter estimation bias and coverage.

Mean peb values averaged over all model parameters per condition are displayed in Figure 7.6.
It is apparent that the mono-construct conditions show considerable average bias for a large number of conditions. These bias values should be interpreted with caution, as they are most probably due to a lack of convergence in the respective conditions (also see Figure 7.5).
On average, bias was higher in the high consistency condition as compared to the low consistency condition, with peb values for high consistency conditions exceeding the cut-off of 10% in almost all cases, including the multi-construct conditions that did not exhibit convergence problems. In those low consistency conditions that did not exhibit convergence problems, parameters were, on average, estimated with acceptable accuracy, except for the conditions with small samples on both levels.
Because of variability in bias across parameter classes, accuracy of parameter estimation is described in greater detail below.
Threshold parameters of mono-construct and multi-construct models were estimated accurately (peb $< 10\%$) in all informative prior conditions and in all of the diffuse prior conditions that did not exhibit convergence problems (see Figure B 19 in Appendix B.2 and B 38 in Appendix B.3 as well as Figure 7.5).

Mono-construct models. Biases per parameter, mean coverage, peb, mean MSE and empirical *SD* values for the parameters in the mono-construct LST-Com GRM are displayed in Figures B 1 - B 19 in Appendix B.2.
As was to be expected due to mentioned convergence problems, bias values were large and coverage low for most of the parameters in the high consistency condition with diffuse priors. The only exception were conditions with 3 or 4 measurement occasions in combination with large between-level ($nL2 \geq 500$) and within-level samples sizes (nL1 = 20). Using informative priors, bias values could be reduced by large amounts in conditions with small sample sizes, however not to sufficient degrees

to reach acceptable bias levels. In conditions with large sample sizes (nL2 $\geq$ 500, nL1 = 20), the use of informative priors yielded acceptable levels for most parameters, with unique and common method trait variances still exhibiting coverage below and bias above the cut-off values.

In the low consistency condition with diffuse priors, bias and coverage yielded acceptable values for all parameters in conditions with a combination of nL2 $\geq$ 500, nL1 $\geq$ 5 and l $\geq$ 3, with the exception of the combination nL2 = 500 with nL1=5. For conditions with two measurement occasions (l = 2) comparable bias and coverage values were observed except for some parameters that needed larger sample sizes, e.g. the common method trait variances.

Setting informative priors on the loading parameters reduced estimation bias and increased coverage rates of the respective parameters, the effects of prior settings on estimation accuracy being largest in conditions with few observations (on the between- and the within-level) and minimal to zero in conditions with large sample sizes. Setting informative priors on loading and threshold parameters did not only decrease bias of the respective parameters, but also, albeit to lesser degrees, of variance and covariance estimates in small sample size conditions. Estimates in large sample sizes were not substantially affected by the prior settings, and bias values in small samples did not reach acceptable levels with informative priors for most parameters. Hence, conditions that yielded acceptable bias values for each parameter class are the same for the diffuse and informative prior conditions in the low consistency models.

Multi-construct models. Biases per parameter, mean coverage, peb, mean MSE and empirical *SD* values for the parameters in the multi-construct LST-Com GRM are displayed in Figures B 20 - B 38 in Appendix B.3.

In the high consistency conditions, parameters (loadings, variances and covariances) of the latent state residual variables and latent trait variables were estimated with high accuracy except for some small sample size conditions (nL2 = 250, nL1 = 2, l = 2) with both informative and diffuse priors. However, unique and common method state loadings and variances exhibited considerable bias and extremely low coverage values in almost all high consistency conditions for both informative and diffuse priors. Biases of method trait and method state residual covariances were lower in absolute value.

In the low consistency conditions, parameters (loadings, variances and covariances) of the latent state residual variables and latent trait variables were estimated with high accuracy for conditions with nL2 $\geq$ 500 and nL1 $\geq$ 5 with both informative and diffuse priors. Also common method trait and unique method trait variances and covariances showed good estimation accuracy in the same conditions. For conditions with only two measurement occasions (l = 2), peb values were slightly higher as compared to the l = 3 and l = 4 conditions for these parameters, requiring nL1 $>$ 5 for the unique method trait variances and nL1 $>$ 10 or nL2 $>$ 500 for the state loadings and variances with diffuse priors, while coverage values were still in the desired range. Unique method state residual and common method state residual loadings and covariances were estimated with good degrees of accuracy as judged by absolute bias values, peb, MSE and coverage in conditions with nL2 $\geq$ 500 and nL1 $\geq$ 5. Parameters showing the highest levels of estimation inaccuracy in these sample sizes of the low consistency conditions were the unique and common method state residual variances ($0.047 < \text{peb} < 0.376$ and $0.081 < \text{peb} < 0.289$, respectively, and $89.4\% < \text{coverage} < 94.8\%$ and $88.8\% < \text{coverage} < 94.8\%$, respectively). Note, however, that absolute bias values were small, ranging from 0.006 to 0.076 and 0.009 to 0.039 for the unique and common method state residual variances, respectively.

As in the mono-construct models, informative priors led to better estimation accuracy primarily in conditions with few observations, did, however, not substantially change required sample sizes.

Standard error bias.

Mean seb values averaged over all model parameters per condition are displayed in Figure 7.6.

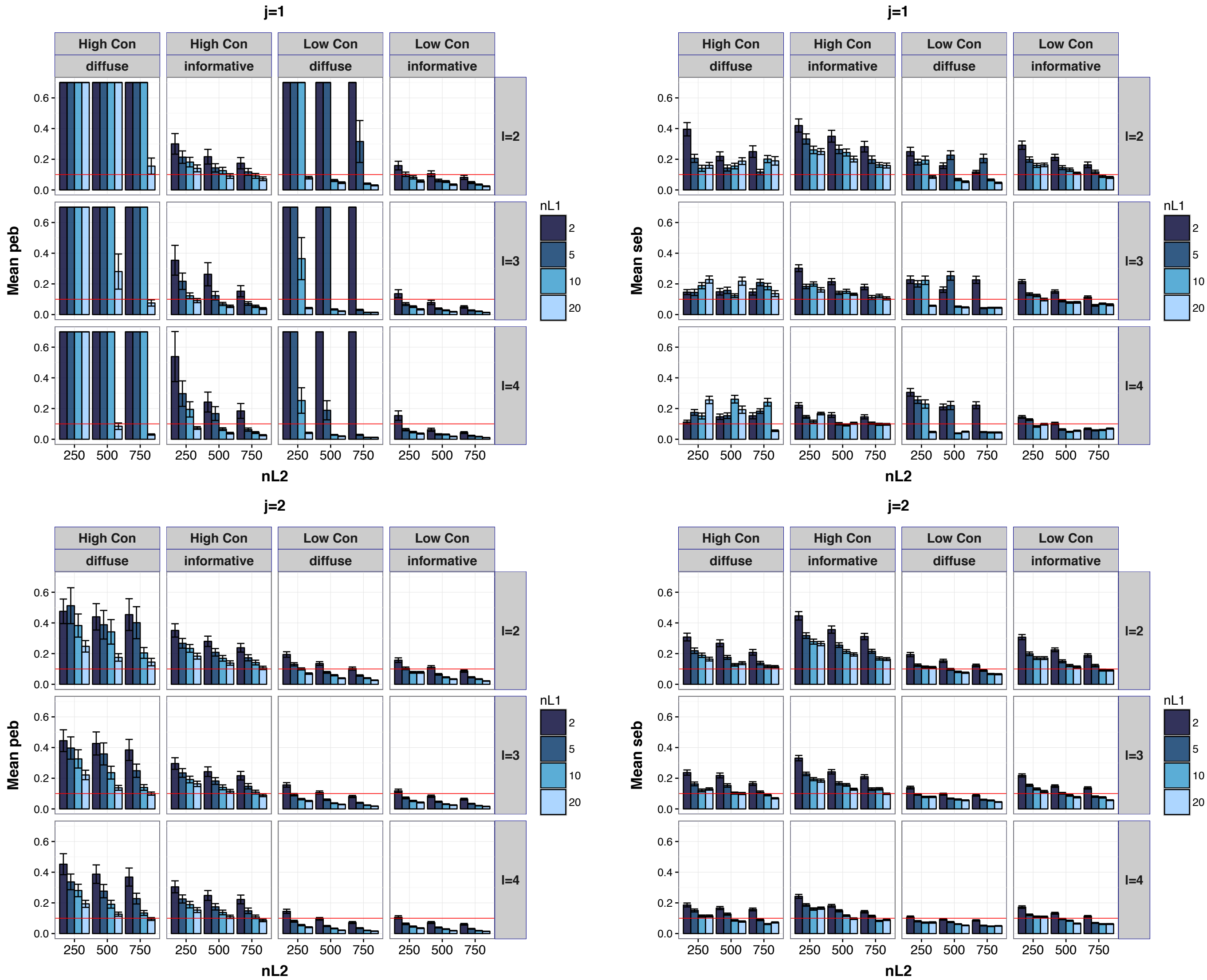

Figure 7.6: Mean parameter estimation bias (peb; left hand side) and standard error bias (seb; right hand side) values averaged over all parameters per condition of the mono-construct model ($j = 1$; upper panel) and the multi-construct model ($j = 2$; lower panel) in the LST-Com GRM simulation study. Error bars represent standard errors. Bars are cut-off at a value of 0.7 to enhance readability of the plots in the relevant range. Diffuse: diffuse prior condition; High Con: high consistency condition; informative: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

In the multi-construct condition, average standard error bias was higher in the informative prior conditions as compared to the diffuse prior conditions, with seb values for the informative prior conditions with few observations mostly exceeding the cut-off of 10%. On average, standard error bias is acceptable in the diffuse prior conditions. In the mono-construct conditions, there is no clear pattern of seb values over conditions, which might be explained by empirical *SD*s and posterior *SD*s being highly affected by non-convergence of some parameters in these conditions.

Differences between empirical and posterior *SD*s per parameter as well as mean empirical *SD*s for the parameters in the LST-Com GRM are displayed in Figures B 1 - B 19 for the mono-construct condition in Appendix B.2, and Figures B 20 - B 38 for the multi-construct condition in Appendix B.3.

As theoretically expected, empirical *SD*s tend to decrease with an increasing number of observations on both the within- and between-level for all parameter types and conditions. Informative priors lead to smaller empirical *SD*s primarily in conditions with few within-level observations and for the parameters they were set on (loading parameters) or parameters that did not converge in the diffuse prior conditions. High empirical *SD*s in the high consistency diffuse prior conditions mirror the non-convergence of some parameters (e.g., the *UMS* and *CMS* variances) in these conditions as reported in section 7.6.2. Dot plots of standard error bias for the loading parameters show that posterior *SD*s mainly overestimate empirical *SD*s for the respective parameters when informative priors are used. A similar pattern can be observed for the standard error bias of the state residual variances, where the tendency of overestimation subsides with an increasing number of measurement occasions. In contrast, posterior *SD*s of the remaining variance and covariance parameters tend to overestimate empirical *SD*s with both diffuse and informative priors, bias levels decreasing with increasing sample sizes. When ignoring conditions with high levels of non-convergence (entailing large empirical *SD*s), no systematic pattern of differences between informative and diffuse prior conditions can be observed for these parameters.

7.6.3 Summary and conclusion

The simulation study shows that the LST-Com GRM can be accurately estimated with more than 250 between-level and at least 5 within-level observations with moderate degrees of convergent validity of the methods often found in empirical applications (consistency around .3). Estimation accuracy in the LST-Com GRM is better for models with more than two measurement occasions, and convergence problems and bias are smaller in multi-construct as compared to mono-construct models. Thus, it seems recommendable to include at least three measurement occasions and more than one construct in the model designs. Note that in MTMM analysis, the recommendation to include at least two constructs is actually not an additional requirement, as multi-trait analyses should include more than one construct by definition.

The simulation study reveals that the applicability of the LST-Com GRM reaches its limits in cases of high convergent validity as well as in mono-construct models with low sample sizes. As discussed for the LS-Com GRM, high convergent validity is rarely encountered in practice, as the high consistency condition was included to investigate an upper bound of bias that is to be expected.

The use of weakly informative priors on loading and threshold parameters in the LST-Com GRM reduced convergence problems in these conditions to considerable degrees, while sample sizes required in order to yield accurate parameter estimates could not be reduced with this degree of prior informativeness. While bias of the loading parameters could be reduced to acceptable levels, bias of parameters that did not receive informative priors (variances and covariances) stayed high in low sample size conditions. Consequently, setting additional informative priors on variance and covariance parameters might enhance the applicability of the models with smaller sample sizes or higher degrees of convergent validity. Furthermore, it is probable that increasing the degree of prior informativeness (using lower prior variances) would reduce the sample sizes needed to reach acceptably low parameter biases. However, setting priors and variance and covariance parameters in complex models is challenging and the use of highly informative priors involves the risk of detrimental effects on parameter estimates in case the prior's locations were not accurate (see, e.g., Holtmann et al., 2016).

7.7 Monte Carlo simulation LGC-Com GRM

7.7.1 Simulation design

The model used in the simulation study is the LGC-Com GRM with indicator-specific intercept and slope variables and common state residual variables, depicted in Figure 7.7. The simulated model

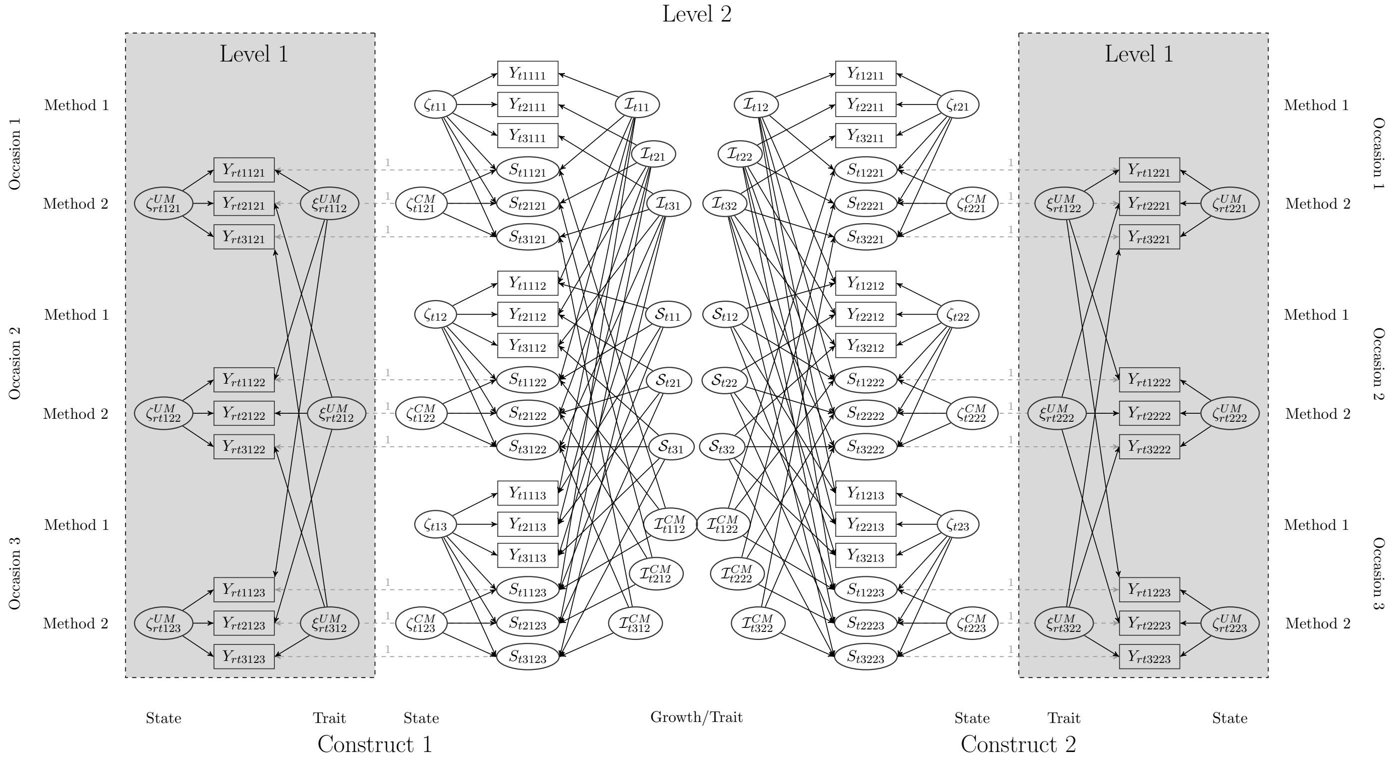

Figure 7.7: Path diagram of the Latent-Growth-Curve-Com graded response model with indicator-specific latent intercept ($\mathcal{I}_{tij}$) and slope ($\mathcal{S}_{tij}$) as well as method intercept ($\mathcal{I}_{tjkl}^{CM}$) variables, and with common latent state residual variables ζ_{tjl}. The model includes one structurally different method and one set of interchangeable methods on three measurement occasions for two constructs. Method 1 is selected as reference method. Note that for illustration purposes, the path diagram is depicted for the observed variables $Y_{(r)tijkl}$, which are probabilistically linked to the latent variables by a probit link. For convenience, the constant indicator $k = 1$ or $k = 2$ has been dropped from the latent trait intercept variables ($\mathcal{I}_{tij}$=$\mathcal{I}_{tij1}$), latent slope variables ($\mathcal{S}_{tij}$=$\mathcal{S}_{tij1}$) and the latent state residual variables (ζ_{tjl}=ζ_{tj1l}). For the sake of clarity, correlations between latent variables and loading parameters are omitted. Note that loading parameters of the latent intercept and slope variables are restricted in order to model a linear growth trajectory. Correlations that are not permissible in the depicted LGC-Com GRM are all correlations between any (method) intercept variable $\mathcal{I}$ or (method) slope variable $\mathcal{S}$ and any state residual (method) variable ζ, correlations between the latent intercept and the latent intercept (common) method variables of the same construct j and indicator i, correlations between the latent slope and the latent slope (common) method variables of the same construct j and indicator i, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. CM: common method; M: method; S: latent state variable; $\mathcal{S}$: latent slope variable; UM: unique method; $\mathcal{I}$: latent intercept variable; ξ: latent trait variable; Y_{rtijkl}: observed variable for the rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

assumed trait change only in the reference-method, that is, the model did not include method slope variables (but only method intercept variables). Note that assuming no trait change in the non-reference methods that is not explained by the reference-method trait change, means that the within-level part of the model corresponds to the LST-Com GRM. Given that the specification of a latent growth curve model requires at least three measurement occasions, the conditions of measurement occasions were reduced to 3 and 4 in the simulation. Additionally, as the LGC-Com GRM is even more complex than the LST-Com GRM, the sample sizes simulated on the between-level were increased to 400, 600, and 800. Five aspects were manipulated in the simulation: (a) the number of constructs (1 or 2), (b) the degree of convergent validity (high vs. low), (c) the number of measurement occasions (3 or 4), (d) the number of level-1 units (i.e., raters per target: nL1 = 2, 5, 10, and 20), (e) the number of level-2 units (i.e., targets: nL2 = 400, 600, and 800), and (f) the amount of prior information used in the estimation (diffuse vs. informative priors), resulting in a 2x2x2x4x3x2 design. In total, the simulation study was comprised of 192 conditions and was simulated with 200 replications per condition.

The population model parameters equal those in the LST-Com GRM simulation study, with latent trait variances being split into 92.16% of trait intercept variance and 7.84% of slope variance. Growth trajectories were modeled to be linear. Population values for the different variance components in the two consistency conditions on the first measurement occasion correspond to those of the LST-Com GRM and can be found in Table 7.2. Population values of parameters differing from the LST-Com GRM are given in Table C 1. All models were specified assuming strong measurement invariance for all items over measurement occasions (Meredith, 1993; Meredith & Teresi, 2006).

Extensive pre-analyses were used to determine an appropriate number of iterations, burn-in samples and thinning. Trace-plots of MCMC samples and PSR values were used to examine convergence. Based on these analyses, Bayesian estimation of the LGC-Com GRM was conducted running 2 MCMC chains and using a thinning factor of 5. The iterations per chain were varied by condition, based on convergence behavior of the chains in the respective condition (increasing iterations for lower sample sizes and higher model complexity) and estimation times (not increasing iterations for conditions where not necessary). Iterations were fixed to the following values: (a) in the informative prior conditions, 200,000 for conditions with 2 or 5 within-level observations and 150,000 for conditions with 10 or 20 within-level observations; (b) in the diffuse prior conditions, 300,000 for conditions with 2 or 5 within-level observations, 250,000 for conditions with 10 within-level observations, and 200,000 for conditions with 20 within-level observations. The number of iterations divided by 2 were run as burn-in iterations and as iterations after burn-in per chain, using only every fifth iteration as a sample for the posterior distribution.

At the time of analysis, the simulation was not yet completed. Estimation times per replication lay between 1.2 hours for the smallest model (one construct, three measurement occasions, nL2 = 400, nL1=2, informative priors) and 43.5 hours for the largest model available at the time of analysis (two constructs, four measurement occasions, nL2 = 600, nL1 = 20, diffuse priors). Simulating on 15 computers, this resulted in an approximate estimation time of 13 months for the part of the simulation available at the current time point. That is, simulation times were considerable and exceeded the time that was expected based on the simulation times of the LS-Com and LST-Com GRMs. Within the months of simulation, an electrical power outage in the whole university building interrupted the simulations. Due to this interruption, information on PSR values and parameter coverage for the affected conditions are only available for the replications simulated after the power outage. Also,

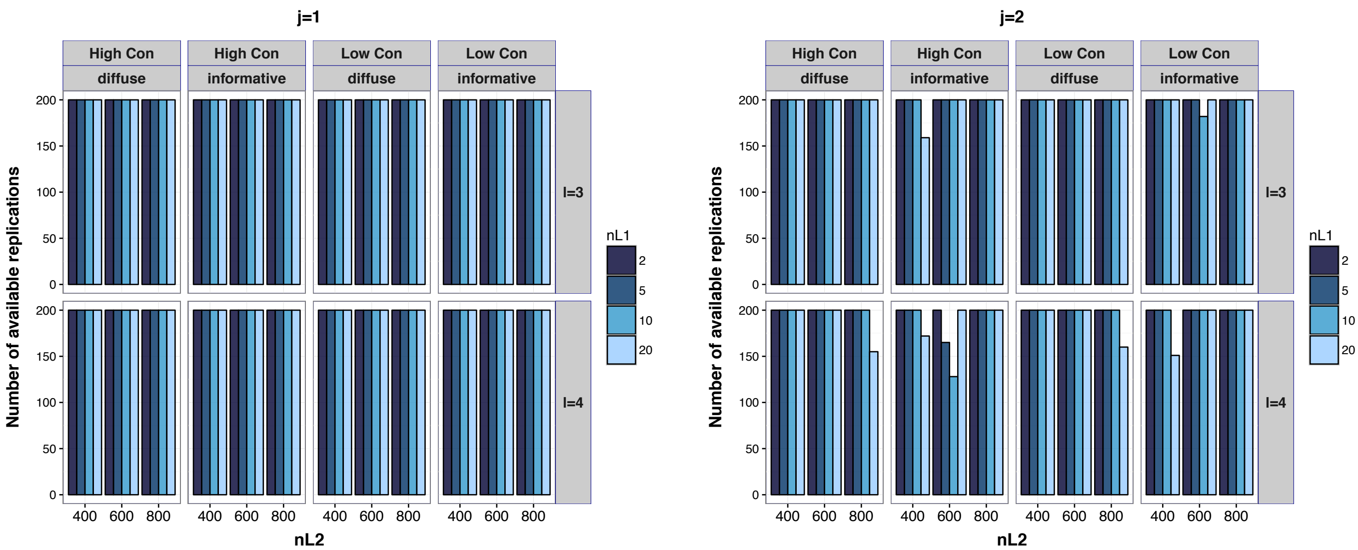

Figure 7.8: Number of available replications for the mono-construct model ($j = 1$; left hand side) and the multi-construct model ($j = 2$; right hand side) in the LGC-Com GRM simulation study. Diffuse: diffuse prior condition; High Con: high consistency condition; informative: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

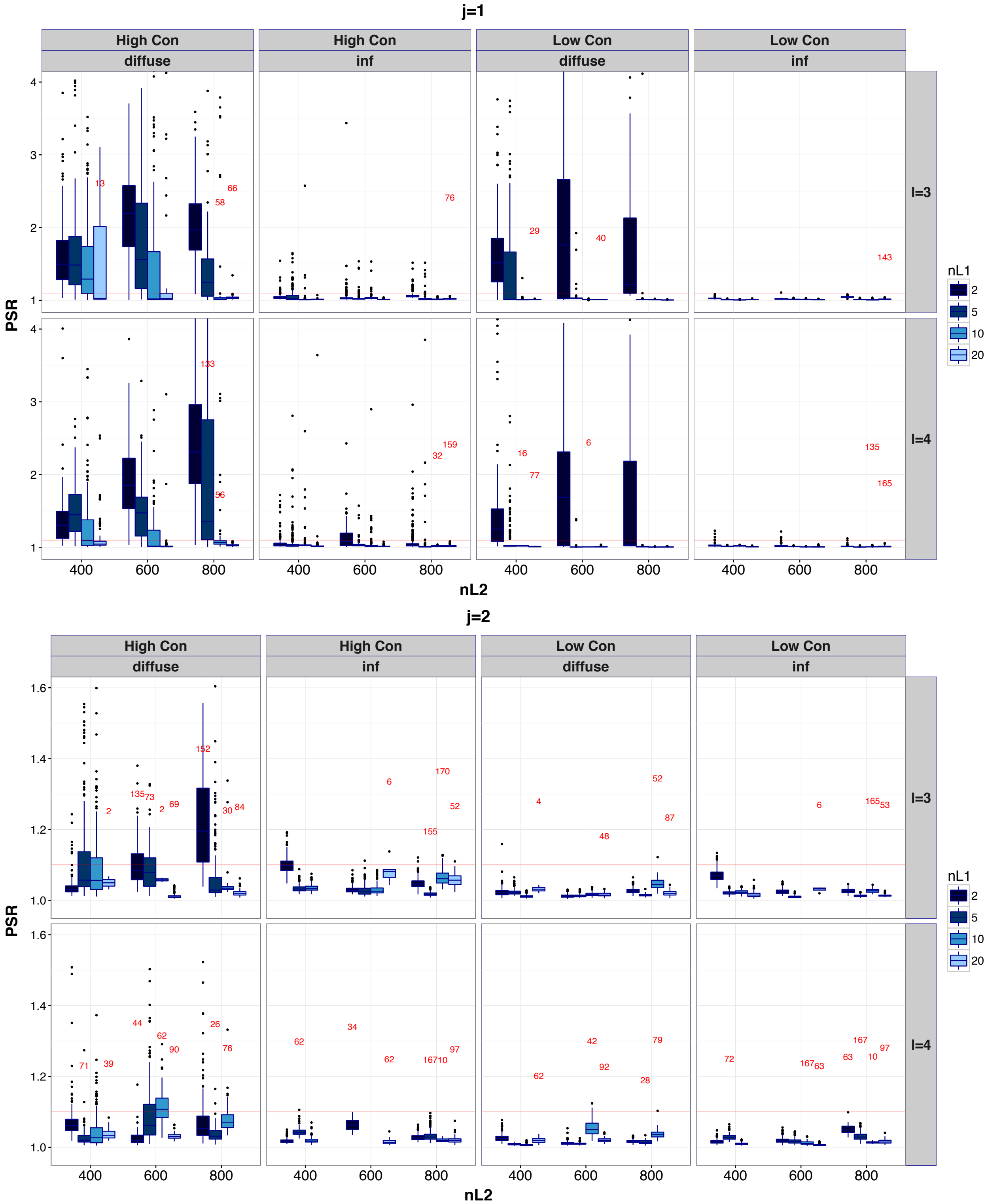

Figure 7.9: Boxplots of PSR values at the last iteration over all 200 replications per condition in the LGC-Com GRM simulation study, for the mono-construct condition ($j = 1$) and the multi-construct condition ($j = 2$). Numbers in red indicate the number of replications for which PSR values were available in the respective condition. No number indicates that all 200 replications provided PSR values in the respective condition. Note that y-axes are only displayed up to a maximum value of 4.2 and 1.65 to enhance readability of the plots in the relevant range. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations; PSR: Potential scale reduction.

they are only yet available for conditions with a completed simulation. The number of replications for which PSR values are available in each condition are displayed in the respective figure showing the distribution of the PSR values (Figure 7.9). Information on parameter and standard error estimates as well as model fit (PPP) are available for all replications and conditions simulated up to the present time point. Figure 7.8 gives an overview of the number of replications available at this time point and included in the current analyses.

7.7.2 Results

Convergence

Figure 7.9 displays the distributions of the PSR values at the last iteration over the different replications per condition. In the mono-construct condition ($j = 1$), the estimation exhibited convergence problems under high consistency with diffuse priors, with few exceptions in conditions with large sample sizes on both levels. Convergence problems seem to disappear to large degrees when using informative priors in the high consistency condition. In the low consistency condition of the mono-construct models, convergence problems were restricted to conditions with only 2 observations on the within-level or with 5 observations on the within-level in combination with few observations on the between-level. Parameters that showed by far the most convergence problems in these conditions were the loading parameters of the slope factors on the non-reference method items. Note that these parameters were specified as regression parameters of the regression of a slope factor of the non-reference method indicators on a slope factor of the reference method indicators, the residual of which was fixed to zero. The factor loadings of these factors were fixed by the definition of the growth curve model. This reparameterization is mathematically equivalent to the original parameterization, was, however, necessary due to limitations in the model specifications for Bayesian models in Mplus. Besides the slope regression parameters, parameters frequently involved in convergence problems comprised state residual loadings as well as intercept loadings on the interchangeable informant report items. Results concerning the respective parameters and conditions should be interpreted with caution.

In the multi-construct condition convergence was mainly an issue in the high consistency condition with diffuse priors. Parameters with highest PSR values in the multi-construct conditions were also primarily the slope regression coefficients. Most of the other conditions of the multi-construct models (low consistency, informative priors) showed good convergence rates as judged by PSR and trace plots.

Model fit

PPP values were $< .9$ in all replications over all conditions. PPP values smaller than .1 were observed in 22.1% of the mono-construct and 27.1% of the multi-construct conditions. These were primarily high-consistency conditions with large within-level sample sizes. 0.35% to 4.5% of the replications in the respective conditions had PPP values smaller than .1. Recall that a PPP $< .1$ or $< .05$ would indicate poor model fit, according to the recommendations given by B. Muthén and Asparouhov (2012).

PPP values between .1 and .2 occurred in 80.0% of the mono-construct and 74.0% of the multi-construct conditions. 0.5% to 34.3% of the replications per condition showed values in this range, with higher percentages in the high-consistency than the low consistency condition. Between 3.0% and 52.3% and between 3.1% and 55.6% of the PPP values per condition lay between .2 and .3 in the mono-construct and multi-construct models, respectively, with higher percentages in the high consistency conditions. In the mono-construct models, a minimum of 25.0% up to a maximum of

97.0%, and in the multi-construct models, a minimum of 20.1% up to a maximum of 96.9% of the PPP values per condition fell into the range between .3 and .7.

Parameter estimation bias and coverage.

Mean peb values averaged over all model parameters per condition are displayed in Figure 7.10.
It is apparent that the mono-construct conditions show considerable average bias for a large number of diffuse prior conditions. These bias values should be interpreted with caution, as they are most probably due to a lack of convergence in the respective conditions (see Figure 7.9).
On average, bias was higher in the high consistency condition as compared to the low consistency condition, with peb values for high consistency conditions exceeding the cut-off of 10% in all cases. Also in the low consistency condition bias values exceeded the cut-off of 10% in almost all cases, except for conditions with four measurement occasions and large sample sizes on both levels.

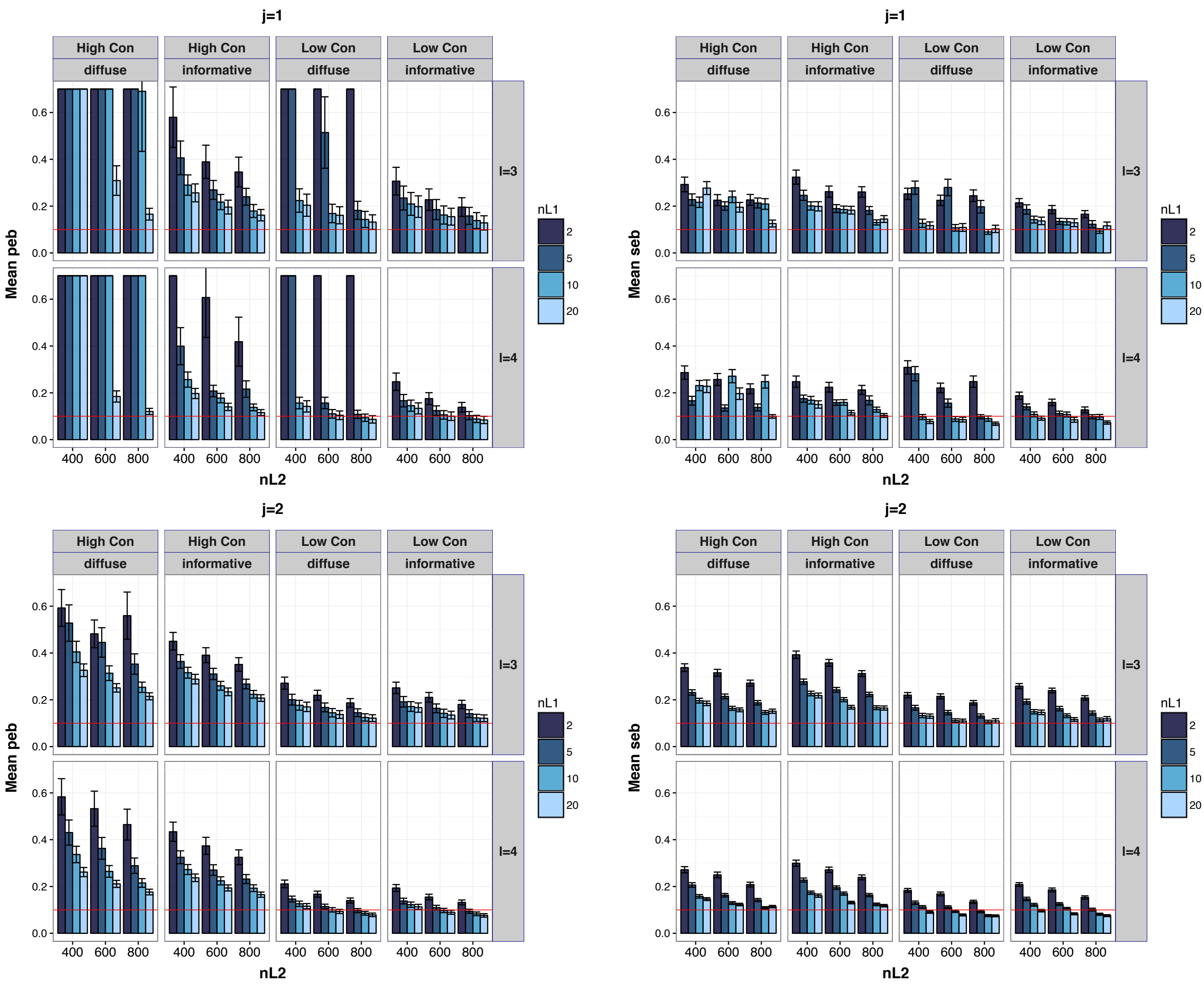

Figure 7.10: Mean parameter estimate bias (peb; left hand side) and standard error bias (seb; right hand side) values averaged over all parameters per condition of the mono-construct model (j = 1; upper panel) and the multi-construct model (j = 2; lower panel) in the LGC-Com GRM simulation study. Error bars represent standard errors. Bars are cut-off at a value of 0.7 to enhance readability of the plots in the relevant range. Diffuse: diffuse prior condition; High Con: high consistency condition; informative: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

As there was a considerable variability in bias across parameter classes, accuracy of parameter estimation is described in greater detail below.

Mono-construct models. Biases per parameter, mean coverage, peb, mean MSE and empirical *SD* values for the parameters in the mono-construct LGC-Com GRM are displayed in Figures C 1 - C 23 in Appendix C.2.
As was to be expected due to mentioned convergence problems, bias values were large and coverage low for most of the parameters in the high consistency condition with diffuse priors. The only exception were conditions with large between-level (nL2 $\geq$ 600) in combination with large within-level samples sizes (nL1 = 20). Using informative priors, bias values could be reduced to acceptable bias levels for loading and threshold parameters in large sample sizes. However, peb values of variance and covariance estimates in the high consistency conditions remained high under the estimation with informative priors. The parameters exhibiting the largest bias problems (high peb values and low coverage) are the slope loadings and variances as well as the unique method state variances.
In the low consistency condition with diffuse priors, bias and coverage yielded acceptable values for most of the parameters in conditions with a combination of nL2 $\geq$ 600, nL1 $\geq$ 10. Parameters still showing bias for these sample sizes are the state residual variances, the slope loadings, the slope variances as well as the slope covariances. Note, that for instance for the slope covariances, peb values are large even in large sample sizes, but MSE values are small and coverage is good. This might be due to the small population values of the slope covariances in the data generation, which was set to 0.031. Similarly, the slope variances show small absolute biases and MSEs, while peb values are large and coverage is insufficient, which again might be due to small population values as well as small empirical SDs. Latent slope means showed acceptable peb values and small MSE, however low coverage values.
Setting informative priors on the loading parameters reduced estimation bias and increased coverage rates of the respective parameters, the effects of prior settings on estimation accuracy being largest in conditions with few observations (on the between- and the within-level) and minimal to zero in conditions with large sample sizes. Estimates in large sample sizes were not substantially affected by the prior settings, and bias values in small samples did not reach acceptable levels with informative priors for most parameters.

Multi-construct models. Biases per parameter, mean coverage, peb, mean MSE and empirical *SD* values for the parameters in the multi-construct LGC-Com GRM are displayed in Figures C 24 - C 46 in Appendix C.3.
In the high consistency conditions, parameters (loadings, variances and covariances) of the latent state residual variables and latent intercept variables were estimated accurately except for some small sample size conditions, with both informative and diffuse priors. Slope loadings showed some bias and low coverage in all high consistency conditions and slope variances showed a similar bias pattern as in the mono-construct models, i.e., low coverage and high peb values while absolute bias and MSE values are small. However, also unique and common method state loadings and variances exhibited considerable bias and extremely low coverage values in almost all high consistency conditions for both informative and diffuse priors. Biases of method state residual covariances were lower in absolute value.
In the low consistency conditions, parameters (loadings, variances and covariances) of the latent state residual variables and latent trait variables were estimated with high accuracy with both informative and diffuse priors. Also common method intercept and unique method trait variances and covariances showed good estimation accuracy in conditions with nL1 $\geq$ 5. Unique method state residual and common method state residual loadings were estimated with good degrees of accuracy as judged by absolute bias values, peb, MSE and coverage in conditions with nL1 $\geq$ 5. Covariances of the same

factors required larger sample sizes, preferably nL2 $\geq$ 600 and nL1 $\geq$ 10, to reduce peb values and absolute biases, while coverage was acceptable and MSE values small under all sample sizes.
Parameters showing estimation inaccuracy in the low consistency conditions were the unique and common method state residual variances. Note, however, that for nL1 $\geq$ 5, absolute bias values were small, ranging from 0.007 to 0.076 and 0.012 to 0.049 for the unique and common method state residual variances, respectively.
Parameters showing the highest peb values and low coverages in the multi-construct low consistency conditions are the slope loadings and slope variances. Note that the high peb values for the slope variance parameters might be due to their small population values (i.e., 0.038), as absolute bias and MSE values are quite small. However, also coverage values indicated that slope loading and variance parameters were not estimated with sufficient accuracy in any of the conditions.
Setting informative priors on loading and threshold parameters did not substantially change any bias or coverage results in the multi-construct low consistency models. Increasing the number of measurement occasions from three to four did have a minimal positive impact on some of the parameters' estimation accuracy, e.g., the intercept and slope loadings and variances. However, differences in estimation bias were rather small.

Standard error bias.

Mean seb values averaged over all model parameters per condition are displayed in Figure 7.10.
On average, there was a minimal tendency of larger seb values in the informative as compared to the diffuse prior conditions. Seb values in the high consistency conditions mostly exceeded the cut-off of 10%, while, on average, standard error bias is acceptable in the low consistency conditions with large sample sizes (e.g., nL1 $\geq$ 10).
Patterns of seb values over conditions are not systematic, which might be explained by empirical *SD*s and posterior *SD*s being highly affected by non-convergence of some parameters in these conditions. Differences between empirical and posterior *SD*s per parameter as well as mean empirical *SD*s for the parameters in the LGC-Com GRM are displayed in Figures C 1 - C 23 for the mono-construct condition in Appendix C.2, and Figures C 24 - C 46 for the multi-construct condition in Appendix C.3.
The main pattern observable in the empirical *SD*s mirrors the non-convergence of some parameters (e.g., the slope loadings) as reported in section 7.7.2. That is, high empirical *SD*s occur in the conditions that showed convergence problems, i.e., mainly mono-construct high consistency and diffuse prior conditions. In these conditions, the use of informative priors lead to smaller empirical *SD*s, in the same way as it improved convergence. As theoretically expected, empirical *SD*s tend to decrease with an increasing number of observations on both the within- and between-level for all parameter types and conditions.
Dot plots of standard error bias show that posterior *SD*s of the state residual, unique method state and common method state loading parameters mainly overestimate empirical *SD*s if informative priors are set on these loadings. No systematic pattern of differences between informative and diffuse prior conditions can be observed for the remaining parameters. In general, standard error bias is small in those conditions that did not show convergence problems, i.e., standard error bias is < 0.04 in absolute value.

7.7.3 Summary and conclusion

The simulation study shows that the estimation of the LGC-Com GRM encounters convergence problems in high consistency conditions or low consistency conditions with small sample sizes and diffuse priors, primarily with respect to the slope factor loadings. Setting informative priors on the loading parameters effectively reduced convergence problems in small samples, does, however, not yield acceptable bias levels in these conditions.

The applicability of the LGC-Com GRM reaches its limits in cases of high convergent validity. However, as discussed previously, high convergent validity is rarely encountered in practice.

Applicability of the LGC-Com GRM seems most realistic for models with moderate degrees of convergent validity and sample sizes of at least 600 between-level and 10 within-level observations. Moderate degrees of convergent validity as simulated in the low consistency condition are often found in empirical applications. However, even in cases of low consistency and sample sizes as large as nL2 = 600 and nL1 = 10, some parameters of the LGC-Com GRM show unacceptable bias and coverage levels. These parameters are the slope loadings and slope variances. Note that these parameters show a pattern of bias and coverage similar to the method state residual variables in the high consistency condition. As the slope variances were simulated to be rather small, convergence problems and bias encountered for the slope loadings and variances in the LGC-Com GRM can most probably be traced back to the same underlying mechanism as in case of the problems for the method state variables under high consistency. This is, if factor variances are very small, there is also little information in the data in order to estimate these factors, empirical identification of the slope factors becomes difficult and estimation runs into problems. A similar observation was made by Bishop et al. (2015) in a simulation on the ISGM, where growth factor variances were most biased when (time-) consistency was low, i.e. when most of the true score variance was due to state variability processes. They concluded that weakly defined growth factors (i.e., low time consistency) can be problematic, especially when other aspects, such as the number of measurement occasions or sample size, are suboptimal as well (Bishop et al., 2015). Similarly, the power to detect significant slope variances in LGC models has been found to not only be a function of sample size and number of measurement occasions but also of growth curve reliability, which corresponds to time consistency in multiple-indicator models (see also Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008; von Oertzen et al., 2010).

Thus, it is to be expected that the LGC-Com GRM can be accurately estimated and slope variable parameters run into less convergence problems if slope factor variances are larger, i.e., if inter-individual differences in intra-individual change are larger.

Nevertheless, it is a realistic phenomenon that slope factor variances are small and it is often found in empirical applications that factors need to be excluded from the model as there is no remaining variance present in the data they could capture.

In their simulation study on indicator-specific growth models (ISGM) Bishop et al. (2015) found that the parameters most prone to estimation problems were the slope variances, the intercept correlations and the slope correlations. Based on their findings they recommended using sample sizes of at least 300 and at least four time points for the ISGM (Bishop et al., 2015). In contrast to their simulation study, the current simulation did not find substantial estimation advantages if four instead of three measurement occasions were included in the model. The only model parameters showing a relevant improvement in estimation accuracy when increasing the number of measurement occasions from three to four were the slope factor parameters (i.e., slope loadings, variances and covariances). Note, however, that bias was already very small and coverage good for the slope covariances in case of three measurement occasions. The slope loadings and variances, in contrast, had poor coverage also with four measurement occasions. Nevertheless, the observed pattern suggests that bias levels can be expected to decrease and coverage to increase when further increasing the number of measurement occasions. Furthermore, the effect of the number of measurement occasions might be more prominent if there is a larger variance in slope factor values.

The fact that larger samples are required in the LGC-Com GRM as compared to the ISGM by Bishop et al. (2015) is not surprising, as the LGC-Com GRM is comparatively more complex, including a number of additional latent factors due to the MTMM structure as well as a multilevel structure due to the interchangeability of the non-reference method raters.

The use of weakly informative priors on loading and threshold parameters in the LGC-Com GRM reduced convergence problems to considerable degrees, did, however, not reduce sample sizes required in order to yield accurate estimates for most of the parameters. Furthermore, they did not resolve problems in the estimation of the slope variable parameters.
Increasing prior informativeness or setting additional informative priors on variance and covariance parameters might enhance the applicability of the models with this degree of slope variance, smaller sample sizes or higher degrees of convergent validity. However, setting priors and variance and covariance parameters in complex models is challenging and the use of highly informative priors involves the risk of setting priors with inaccurate locations.
Also, setting more informative priors might not be necessary for accurate parameter estimation in the LGC-Com GRM if inter-individual differences in change are actually larger in the observed sample. However, if slope variances are estimated to be very small and the estimation shows convergence problems for the respective slope parameters, it might indicate that most of the true score variance is due to state variability processes and / or stable (non-changing) inter-individual differences. In this case, researchers should consider to resort to more simple models that are more appropriate for this kind of processes, such as the LST GRM.

7.8 Discussion of the simulation studies

As the results of the simulation studies show, the LS-Com GRM and LST-Com GRM can be accurately estimated if low degrees of convergent validity are present. While the LS-Com GRM requires at least 250 observations on the between-level and more than 2 (better 5) observations on the within-level, the LST-Com GRM requires larger sample sizes of at least 500 between- and 5 within-level observations when more than two measurement occasions are included in the model.
Note that the degree of convergent validity chosen for the high consistency condition is rarely encountered, while low convergent validity is frequent in practical applications (see, e.g., the following empirical example, or Carretero-Dios et al., 2011; Eid et al., 2003, 2008). The high consistency condition provides an upper bound of consistency that would be desirable in multirater studies. That is, the bias values reported for the high consistency condition can be interpreted as an upper bound of bias that could be encountered in applications, with far smaller biases that are to be expected in practice (e.g., those of the more realistic low consistency condition).
Hence, in general, unbiased results are to be expected with these moderate sample sizes. These results are encouraging and indicate that even complex multilevel MTMM IRT models can be applied in a wide range of situations when estimated with Bayesian methods.
However, with only few level-1 units (e.g., 2 observations per cluster) convergence problems become more likely and bias levels increase. Also, in contrast to the continuous indicator models, more observations (approximately 250 (LS-Com GRM) or 500 (LST-Com GRM) instead of 100) are required on the between-level. This result supports previous findings suggesting that categorical models need larger sample sizes on both levels than continuous-indicator SEMs (Depaoli & Clifton, 2015; Holtmann et al., 2016; Lee et al., 2010), as well as results on single-level CFA-MTMM models with ordinal variables reporting a need of at least 250 observations when two to four indicators are used per TMU (Nussbeck et al., 2006). However, given the complexity of the presented models, the required sample sizes can be considered moderate.
In contrast to simulation studies indicating a greater influence of between-level as compared to within-level sample sizes on parameter estimates (Maas & Hox, 2005), the simulation results show that the number of level-1 units has a substantial impact on convergence rates and estimation accuracy of the within-level parameters. This corresponds to the result of the simulation study by Koch et al. (2014) on the continuous indicator counterparts of the GRMs simulated here, as well as the results of

Bayesian estimation of categorical-indicator multilevel SEMs in Holtmann et al. (2016). Furthermore, similar to a result by Depaoli and Clifton (2015) on multilevel categorical SEMs, the number of level-1 units and the degree of convergent validity (or ICCs) had the largest impact on convergence rates.

In the unlikely situation where high degrees of convergent validity are present in the data, a larger number of observations is needed to yield sufficiently accurate parameter estimates and ensure convergence. High degrees of convergent validity are especially problematic in the LST-Com GRM (and LGC-Com GRM) as well as in mono-construct models. The results indicate that in the LS-Com GRM, the use of informative priors on item parameters (such as loadings and thresholds) may effectively reduce biases in situations where an increase of observations is not feasible. That is, the possibility to incorporate informative prior information in the estimation process further increases the applicability of the model in rather small samples. By using weakly informative priors, researchers may also lower the risk of improper solutions such as negative variances and thereby increase convergence rates that might otherwise cause problems in this kind of complex models. The use of accurate informative priors does not only decrease bias for the parameters that were given informative priors (e.g., the loadings), but also, albeit to smaller degrees, for the remaining parameters (e.g., the variances). The degree of prior information chosen in the present studies, (a prior variance of 0.1) did, however, not affect the results in cases where sufficient information in terms of observations was provided. This is in line with theoretical considerations as well as previous results reporting a decrease of the prior's influence with increasing sample size (Asparouhov & Muthén, 2010b; Lee et al., 2010).

In the LST-Com GRM the use of informative priors effectively increased convergence rates and decreased bias, however, the degree of prior information used is not sufficient in order to estimate the models with smaller sample sizes than those recommended for diffuse prior settings.

Interestingly, standard error bias increased with the use of informative priors, average posterior *SD*s mostly overestimating empirical *SD*s for the respective parameters. This finding is in accordance with previous results on Bayesian posterior *SD*s (Lee & Song, 2004; Lee et al., 2010). Although posterior *SD*s do not play a role as important for hypothesis testing as standard errors do in classical statistical approaches, high posterior *SD*s might be indicative of too large variation in the posterior distribution. Coverage values, however, did not necessarily point into this direction (if observations on the between- and within-level reached the recommended values).

Low rejection rates of the PPP at the 10% and also 20% level have been reported before for PPP values in correctly specified models (Asparouhov & Muthén, 2010b). This observation might stem from the fact that PPP values do in general not have a uniform distribution under the null hypothesis (Gelman, 2013; Hjort, Dahl, & Steinbakk, 2006; Meng, 1994). Whether these low rejection rates of correctly specified models represent a problem or are in fact a desirable property is an ongoing discussion (see, e.g., Gelman, 2013; Hjort et al., 2006, and references therein).

In contrast to the LS-Com GRM and LST-Com GRM, the estimation of the LGC-Com GRM poses more problems. That is, even in case of moderate convergent validity (low consistency condition) and sample sizes as large as nL2 = 600 and nL1 = 10, the slope loading and variance parameters are not estimated with sufficient accuracy, as indicated by bias and coverage levels. It is probable that the estimation problems encountered for the slope parameters are due to small simulated factor variances. That is, with small slope variances there is little information in the data in order to estimate these factors and the estimation runs into problems. It is to be expected that the LGC-Com GRM can be accurately estimated and slope variable parameters run into less convergence problems if slope factor variances are larger, i.e., if inter-individual differences in intra-individual change are larger. However, small slope factor variances are often encountered in practice (see, for instance, the following application study) and therefore a realistic phenomenon.

An improvement in estimation accuracy when increasing the number of measurement occasions from three to four was observed for the slope factor parameters (i.e., slope loadings, variances and covariances). However, the slope loadings and variances had poor coverage also with four measurement occasions. Nevertheless, the observed pattern suggests that bias levels can be expected to decrease and coverage to increase when further increasing the number of measurement occasions. Furthermore, the use of weakly informative priors on loading parameters in the LGC-Com GRM did not resolve problems in the estimation of the slope variable parameters. Increasing prior informativeness or setting additional informative priors on variance parameters might enhance the applicability of the models with this degree of slope variance. However, setting more informative priors might not be necessary for accurate parameter estimation in the LGC-Com GRM if inter-individual differences in change are actually larger in the observed sample. Furthermore, as discussed before, increasing informativeness of priors increases the risk for detrimental effects in case of wrong prior locations. A limitation of the simulation study on the LGC-Com GRM is, that the effect of the amount of slope variance was not tested by including an additional simulation condition with larger slope variance parameters. Including an additional slope variance condition in the present simulation study was considered infeasible due to the large estimation times of the LGC-Com GRM simulation. The hypothesis that estimation problems of the LGC-Com GRM slope parameters might vanish if the slope factor exhibits a larger variance should be investigated in future studies within a smaller design (e.g., including only models with low consistency and sample sizes of 600 between-level observations).

Another limitation of the present simulation studies is the lack of an incorrect prior condition. A detailed investigation of the magnitude of possible detrimental effects of incorrectly specified priors on parameter estimation was considered unfeasible within the scope of the present work. This is due to the estimation times required for the estimation of such complex models by MCMC methods. However, results on the influence of inaccurate priors on estimation accuracy for a multilevel MTMM model for interchangeable raters are reported in Holtmann et al. (2016). The simulation study by Holtmann et al. (2016) showed that when setting weakly informative (prior variance 0.2) inaccurate priors on loading parameters, model parameters were estimated as accurately or even better than with Bayesian estimation with diffuse priors. A similar result has also been reported for the effect of inaccurate priors on growth parameters in Bayesian growth mixture modeling (Depaoli, 2014). However, increasing the informativeness of the inaccurate priors (prior variance of 0.01) was observed to cause considerable detrimental effects on estimation accuracy (Holtmann et al., 2016).
The inclusion of the correctly specified informative prior condition in the present work already provides interesting information. First, the results show that informative priors do not work equally well in all conditions. As can be expected, the informative prior setting does not perform worse than the diffuse one, but it does also not work better in all cases. The present results show at which sample sizes the informative prior does not improve parameter bias as compared to the uninformative case. Accordingly, it reveals at which sample sizes the analysis can be conducted without access to any prior information and still yield parameter estimates as valid as if one had correct prior information available. For example, in the LS-Com GRM, differences between diffuse and informative prior settings became negligible with samples sizes ≥ 500 on the between-level in combination with ≥ 5 on the witihin-level. In the LST-Com GRM multi-construct models, these numbers were comparable in case of three or four measurement occasions and slightly larger in models with two measurement occasions. Second, we can observe the effect the informative priors on the loading parameters exert on the estimation of the remaining parameters in the models (e.g., the variances). The size of this effect cannot be derived theoretically. Furthermore, the simulation results also show the effect of the prior on the estimation of the posterior *SD*s.
Theoretically, it should not be too unrealistic to have information about the location (within a certain range) in which a loading of a second or third indicator of a factor lies, given that the first indicator of that factor was set to 1. Hence, including weakly informative prior information on the location of the loading parameters in the presented models is a realistic option. However, the availability of good a priori information on some model parameters might be scarce in practice. The option to include prior information in the estimation process seems especially interesting for MTMM analysis

if researchers aim to include past findings concerning the convergent and discriminant validity of a particular measure or instrument in future studies. However, the inclusion of prior information on variances and covariances in the model estimation poses a challenge, as Inverse Wishart priors are difficult to handle, especially if variance covariance matrices grow large. Inverse Wishart priors impose a dependency between standard deviations and correlations (Tokuda, Goodrich, Van Mechelen, Gelman, & Tuerlinckx, 2011) and the informativeness of one parameter in the covariance matrix determines the informativeness of other parameters, thereby restricting their prior range (Gelman et al., 2014). These disadvantages render the specification of prior information for variance and covariance parameters extremely challenging and reduce the flexibility of prior settings needed in empirical applications. Alternative prior settings for variance and covariance parameters, such as decomposing covariance matrices into a scale and a correlation matrix (Lewandowski, Kurowicka, & Joe, 2009; Stan Development Team, 2014b) and their applicability to the presented models and MTMM coefficients should be explored in future studies.

Another limitation of the simulation studies is that the results can not necessarily be generalized to categorical observed variables with a different number of categories. Three response categories were chosen in order to present the presumably most problematic case of a polytomous response scale, which excludes the possibility of treating the responses as continuous (using ML estimation). Simulation studies on the performance of WLS estimation for polytomous data have found parameter bias to be unaffected by the number of response categories (Beauducel & Herzberg, 2006; Li, 2016; Yang-Wallentin, Jöreskog, & Luo, 2010) or to decrease when increasing the number of response categories (Flora & Curran, 2004; Forero & Maydeu-Olivares, 2009; Moshagen & Musch, 2014). Similarly, convergence rates were found to increase (Flora & Curran, 2004; Moshagen & Musch, 2014; Rhemtulla et al., 2012; Yang-Wallentin et al., 2010) or be unaffected (Li, 2016) when increasing the number of categories. Although results on the effect of the number of response categories on parameter estimates in Bayesian estimation are scarce, the direction of the effect is assumed to be the same. One study investigating Bayesian methods for confirmatory factor analysis with categorical indicators has found negligible to positive effects of an increased number of response categories on parameter bias (Liang & Yang, 2014). The effect of the number of response categories and its potential interactions with sample size or cell frequencies on estimation accuracy is an interesting question that should be pursued in future research. However, given the complexity of the presented models and the resulting simulation times (several months), an investigation of the effect of the number of categories is beyond the scope of the present simulation studies.

Last but not least, the simulation studies reveal that the peb as a measure of relative bias might not always be the best measure to judge estimation accuracy or compare estimation accuracy between parameters. As the results on some of the parameters, e.g. the slope covariances in the LGC-Com GRM, have shown, small population parameters can produce extremely large peb values even if bias is small. That is, peb values can indicate unacceptable parameter estimation bias (e.g., peb > 1) even if absolute bias is small (e.g., 0.02), coverage values are good and MSE values are small (e.g., 0.005). Therefore, it is recommended to always base the evaluation of estimation results on several different evaluation criteria, including absolute bias and the MSE.

Chapter 8

Application

8.1 Stability of Life Satisfaction and Subjective Happiness and their dynamic interplay

In the following, an application of the LS-Com GRM, LST-Com GRM, and LGC-Com GRM is presented, illustrating the advantages of sampling the model coefficients by Bayesian MCMC methods. This will be done using subjective well-being (SWB) data obtained by self-reports, parent reports and friend reports for recent high-school graduates in Germany. In the last decades, with the advent of positive psychology (Seligman & Csikszentmihalyi, 2000), an emerging body of research started to focus on the experience of positive emotions. Within this research area, an increasing number of studies is dedicated to the investigation of subjective well-being (Pavot & Diener, 2008), its stability over time (e.g., Eid & Diener, 2004; Luhmann et al., 2011) and its relation to other positive or negative emotions, personality traits or life events (e.g., DeNeve & Cooper, 1998; Diener & Chan, 2011; Jovanovic, 2011; Schimmack, Schupp, & Wagner, 2008).
SWB has been conceptualized as consisting of two broad components: an affective and a cognitive component (Diener, 1984; Lucas, Diener, & Suh, 1996; Pavot & Diener, 1993b). The cognitive component has also been described as life satisfaction (Pavot & Diener, 1993b), while the affective component comprises concepts such as positive affect, negative affect or happiness. Although not independent from each other, affective and cognitive components of SWB are considered to provide complementary information: life-satisfaction representing a longer-lasting aspect of SWB relying on conscious values and goals, while the affective component is considered to be of shorter duration and rather unconscious (Eid & Diener, 2004; Luhmann et al., 2011; Pavot & Diener, 1993b). This separation is also supported by differential correlations with potential predictors and outcomes. For instance, major life events show more persistent effects on cognitive than affective SWB (Luhmann, Hofmann, Eid, & Lucas, 2012), and factors such as job status and income have been found to be stronger predictors for cognitive SWB (Diener, Ng, Harter, & Arora, 2010; Schimmack et al., 2008), while affective SWB is more closely related to personality traits (Schimmack et al., 2008; Jovanovic, 2011).
One of the most popular scales in the measurement of life satisfaction is the Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985; Pavot & Diener, 2008). The SWLS was construed as a global measure of life satisfaction, assessing the cognitive component of subjective well-being (Diener et al., 1985; Pavot & Diener, 1993b). Intended to measure more than momentary mood states or transient situational effects, but yet be sensitive to changes in life satisfaction over longer time periods (Pavot & Diener, 1993b), the SWLS is expected to show moderate temporal stability and is considered to be a state variable (Glaesmer, Grande, Braehler, & Roth, 2011). This

expectation is supported by empirical evidence showing considerable temporal stability of SWLS measures (Diener et al., 1985; Pavot & Diener, 1993a), relatively small occasion specificity of SWLS measures (Eid & Diener, 2004), or a greater influence of chronically accessible, stable information than of transient, variable factors on life satisfaction judgments (Luhmann, Hawkley, Eid, & Cacioppo, 2012; Schimmack & Oishi, 2005). Pavot and Diener (1993b) report test-retest stabilities of the SWLS of six different studies ranging from .54 for a 4-years interval, over .62 for a 2-months and .84 for a 1-month interval between measurement time points. A more recent study reported 1-month test-retest correlations of .80 for the SWLS (Steger, Frazier, Oishi, & Kaler, 2006).

Furthermore, convergence between different raters (e.g. self-rating and peer ratings) has been taken as evidence against significant influences of current mood and situational contexts on global well-being measurements. Pavot and Diener (1993a) report correlations between self-report measures and mean peer reports of life-satisfaction ranging from .42 to .49.

Mood and emotions, as aspects of the affective component of SWB, are considered and have been found to be less stable than cognitive parts of SWB (Eid & Diener, 2004). Measures for the affective component of SWB rely on reports of affective states or affective traits. One measure for subjective happiness as a trait is the Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999). Subjective happiness as measured with the SHS was found to be moderately correlated with SWLS measures, with correlation coefficients ranging from .59 to .72 (Lyubomirsky & Lepper, 1999; Schiffrin & Nelson, 2010; Swami et al., 2009; Zhang, Howell, & Stolarski, 2013). In a meta-analysis on longitudinal studies of happiness, Veenhoven (1994) reports happiness to be rather stable in the short term (several months), but not in the long term (years). Lyubomirsky and Lepper (1999) report test-retest stabilities of the SHS from different studies of $r = .61$ (3 weeks), $r = .85$ (1 month interval), $r = .71$ (3 months interval), and $r = .55$ (1 year). Self- and friend rating correlations for the SHS have been found to be as high as .65 (Swami et al., 2009). In a meta-analysis on self-informant agreement in well-being ratings, L. Schneider and Schimmack (2009) found average self-informant correlations of $r = 0.42$, with no difference between life satisfaction and happiness ratings. Instead, their analyses indicated that moderators of self-informant agreement in well-being ratings might be age, with lower self-informant correlations for younger targets (age < 24), as well as the use of single vs. multiple indicators (multiple-item studies reporting higher correlations; L. Schneider & Schimmack, 2009).

The vast majority of studies, however, base stability and rater-consistency estimates on correlations of observed variables. This approach is problematic as measurement error, on the one hand, and instability and rater-inconsistencies, on the other hand, are confounded (Eid & Diener, 2004). The use of correlations based on observed variables might thus lead to biased estimates of stability and consistencies. One of few exceptions is the study by Eid and Diener (2004), applying a multistate-multitrait-multiconstruct model to the SWLS and different mood and affect measures (also see Luhmann et al., 2011). Eid and Diener (2004) found that only 12 to 16% of the variance in life satisfaction is due to occasion-specific influences, while ocassion-specificity for mood reaches 40 to 58%. Furthermore, they report occasion-specific associations between mood and SWLS to be relatively small ($r = 0.13$, $r = 0.23$, and $r = 0.55$ for 3 measurement occasions, respectively; Eid & Diener, 2004).

However, no study so far has investigated inter-rater consistencies or stability of rater-effects for life satisfaction or subjective happiness using latent variable models, that is, taking measurement error into account. Furthermore, most of the approaches used do not allow for the estimation of consistencies or stability coefficients on the item-level, but rely on summary statistics for the whole scale or test-halves. That is, item comparisons or item selection based on item consistencies, stabilities, reliabilities or item difficulties is not possible using these approaches.

The models proposed in this work offer the possibility to (1) study stability of life satisfaction and subjective happiness controlling for measurement error; (2) do this on the item-level and for relatively few items; (3) study rater effects and consistencies controlling for measurement error; (4) investigate the stability of rater effects over time; and (5) study the generalizability of rater effects across the two components of well-being. To investigate the question whether global life satisfaction (as the cognitive component of SWB) or global subjective happiness (as the affective part of SWB) show greater stability over time, which of the two exhibits greater levels of rater-convergence and how stable rater-effects are over time, the LS-Com GRM, LST-Com GRM and LGC-Com GRM[1] are applied to subjective well-being data obtained by self- and friend reports for recent high-school graduates.

8.2 Methods

8.2.1 Participants and procedure

Data analyzed in this article are taken from a longitudinal study assessing change and stability of attachment styles, well-being and loneliness in young adults after high-school graduation[2]. Subjects were approached during their last month before graduation in high-schools in Berlin and Brandenburg. Questionnaires were completed online at four measurement time points over the course of the first year after graduation, with the first measurement occasion taking place two months after graduation. The present analysis includes a subset of the presented scales, measured on occasion 2 (December 2014), 3 (March 2015), and 4 (June 2015), using self-ratings, parent ratings, as well as ratings from several peers (friends) for each target. Note that peer ratings were first collected on measurement occasion 2. The three occasions of measurement used in the subsequent analyses will be referred to as T1, T2 and T3 in the following.
The analyzed sample contains the data of 501 targets, of which 463 participated at T1, 441 at T2 and 430 at T3. For these targets, 366, 342, and 323 parent reports are available for T1, T2, and T3, respectively. Targets were rated by a mean number of 2.22 exchangeable peer raters per target (min = 1, max = 5, mode = 3 raters). The sample was 29.3% male. Targets ranged in age from 17 to 21 years (mean age = 18.22) at T1.

8.2.2 Measures

The two constructs analyzed for the present purpose were life satisfaction and subjective happiness. Overall life satisfaction was measured using the German version of the Satisfaction With Life Scale (SWLS; Glaesmer et al., 2011; Schumacher, 2003). The SWLS is a well-established questionnaire showing good psychometric properties, with an internal consistency of .79 (Vassar, 2008), .87 (Adler & Fagley, 2005; Diener et al., 1985), .86 (Steger et al., 2006), or ranging from 0.79 to 0.89 (Pavot & Diener, 1993b), and $\alpha = .92$ for the German version (Glaesmer et al., 2011). See Pavot and Diener (1993b) and Pavot and Diener (2008) for reviews on the SWLS and Vassar (2008) for a meta-analysis. Happiness was measured with the Subjective Happiness Scale (SHS; Lyubomirsky & Lepper, 1999). The SHS has been shown to be acceptably reliable (with Cronbachs α ranging from .81 to .94; Lyubomirsky & Lepper, 1999). The German version exhibits comparable characteristics to the English version ($\alpha = 0.82$; Swami et al., 2009). For the present study, 5-point rating scales were used, ranging from 1 (does not apply at all) to 5 (applies completely) for the SWLS and from

[1]Note that obviously not all of these models will fit the data equally well and there will be one model that is most appropriate for the data at hand. Nevertheless, in order to provide an illustration of the presented models, of the questions that can be answered with each model and of how the model coefficients are interpreted, all three models are applied to the data.

[2]A different subset of the data has been used in previous publications that pursued different goals using different statistical models (Luhmann, Bohn, Holtmann, Koch, & Eid, 2016).
A subset of the data used in the present analysis, including the same measures but restricted to only two measurement occasions and two methods (self-reports and peer reports) was analyzed with an LS-Com GRM in Holtmann et al. (2017)

Table 8.1: SWLS and SHS items used in the model application

Scale	Item	
SWLS	1	In most ways my life is close to my ideal.
	2	The conditions of my life are excellent.
	3	I am satisfied with my life.
SHS	1	In general, I consider myself: ... not a very happy person / ... a very happy person.
	2	Compared to most of my peers, I consider myself: ... less happy / ... more happy.
	3	Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you? ... Not at all / ... completely.

Note. SWLS: Satisfaction with Life Scale; SHS: Subjective Happiness Scale.

1 to 5 with the respective labelings given in table 8.1 for the SHS. Only the first three items of both scales were included in the analyses. Item formulations are given in Table 8.1. The items excluded from the current analysis were chosen based on substantive reasons. That is, the two SWLS items not included, for instance, are the items ”So far I have gotten the important things I want in life” and ”If I could live my life over, I would change almost nothing”. These items might not be very meaningful in a sample of high-school graduates that are just about to leave their parents home and school environment for the first time in their lives in order to pursue their own interests. Note that item formulations were minimally adapted for the informant peer and parent ratings (replacing the subject ”I” by ”my friend” or ”my child” and changing verb conjugations).

8.2.3 Data analysis

Data were analyzed by applying different variants of the LS-Com, LST-Com and LGC-Com GRMs to the SWLS and SHS self- and informant report measures. The models were fit to the data assuming different levels of factorial invariance. As the only information of model fit currently available for multilevel SEMs with categorical indicators using Bayesian estimation in Mplus (Mplus 7.3) are PPP-values, model comparisons can only be made descriptively in terms of relative fit, but not be tested for signifiant differences using Bayesian estimation. Note also that PPP-values reported by Mplus use the difference between the observed and replicated Chi-Square values, which do not take model complexity into account. To complement the fit information provided for the categorical-indicator models, the models were fit to the data treating the indicators as continuous using the maximum likelihood robust (MLR) estimator in Mplus. The use of the MLR estimator allows for an evaluation of model fit and a comparison of the nested models by χ^2 (difference) tests, fit indices such as the root mean square error of approximation (RMSEA), the comparative fit index (CFI) or the standardized root mean square residual (SRMR). Fit was judged according to the cut-off values given for these indices by Schermelleh-Engel, Moosbrugger, and Müller (2003)[3]. Furthermore, MLR estimation in Mplus provides information criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The results for the continuous-indicator models as estimated with MLR

[3] These are (Schermelleh-Engel et al., 2003): (a) $0 \leq \chi^2/df \leq 2$ good fit; $2 < \chi^2/df \leq 3$ acceptable fit; (b) $0 \leq$ RMSEA $\leq .05$ good fit; $.05 <$ RMSEA $\leq .08$ acceptable fit; (c) $0 \leq$ SRMR $\leq .05$ good fit; $.05 <$ SRMR $\leq .10$ acceptable fit; (d) $.97 \leq$ CFI ≤ 1 good fit; $.95 \leq$ CFI $< .97$ acceptable fit.

are used as a crude indicator of model fit and supplementary information to the Bayesian results. The specific measurement invariance levels that were tested for the different models are described for each model in the respective model section.

Based on these preliminary analyses, the models for categorical indicators were estimated using Bayesian estimation methods with diffuse priors. Posterior parameter estimates were obtained using Bayesian estimation methods with Mplus 7.3 (L. K. Muthén & Muthén, 1998-2012). Trace-plots of MCMC samples were used to examine convergence and determine an appropriate burn-in period. Three MCMC chains were run with a minimum of number of burn-in and sample iterations per chain, using a thinning factor of 10 to reduce autocorrelation. The minimum number of iterations for the LS-Com, LST-Com and LGC-Com GRMs were the following:

- LS-Com GRM: A minimum number of 30,000 burn-in and 30,000 sample iterations per chain, using a thinning factor of 10. That is, at least 300,000 burn-in iterations and 300,000 iterations after burn-in were run, using only every 10-th iteration of the post-burn-in samples for the posterior distributions.
- LST-Com GRM and LGC-Com GRM: A minimum number of 40,000 burn-in and 40,000 sample iterations per chain, using a thinning factor of 10. That is, at least 400,000 burn-in iterations and 400,000 iterations after burn-in were run, using only every 10-th iteration of the post-burn-in samples for the posterior distributions.

Chains stopped running when the PSR reached a value < 1.01 after the minimum number of iterations, or when the PSR had not dropped below 1.05 after a maximum number of 100,000 (burn-in + sample) iterations (no convergence). The mean of the posterior distribution was used as a point estimate and posterior quantiles were used for providing a 95% credibility interval for parameter estimates. Prior specifications were left to the default Mplus diffuse priors (see section 7.3 for details).

As the first category of SWLS item 2 was not chosen by the peers at T3, category 1 and 2 of the respective item were collapsed at T1 and T2 for the peer reports. Similarly, as the first category of SWLS item 1, item 2, and item 3 were not chosen by the parents at at least one measurement occasion, category 1 and 2 of the SWLS items were collapsed for the parent reports.

8.3 LS-Com GRM of Life Satisfaction and Subjective Happiness

In this section results of the application of the LS-Com GRM to the SWLS and SHS self-, parent and peer report measures are presented. In a first step, the LS-Com GRM with no measurement invariance restrictions was estimated with either indicator-specific state or common state variables. These models correspond to the models depicted in Figures 2.1 and 2.2, respectively, with the only difference that the model in this application includes three measurement occasions (instead of two, as depicted in the Figures). Then, different levels of measurement invariance were compared for the chosen model by a stepwise approach. The levels tested were the following three levels: (0) factorial invariance, i.e., no invariance restrictions on the loading or threshold parameters; (1) strong measurement invariance for the reference method (self-report) indicators, i.e., loadings and thresholds set invariant; and invariant loadings of the method factors on the non-reference method indicators; (2) strong measurement invariance, i.e., all factor loadings and threshold parameters are restricted to be equal across measurement occasions for the same indicator i, construct j and method k. Note that this includes the regression parameters of the regression of the non-reference method on the reference-method states,

Table 8.2: Model fit results for the LS-Com GRM of life satisfaction and subjective happiness

MLR estimator (indicators treated as continuous)											
ind.-spec. States	MI	remarks	χ^2	df	RMSEA	CFI	SRMR W	SRMR B	AIC	BIC	BIC adj.
No	0[a]	saddle point, Ψ	3130 .70	1398	.030	.887	.093	.115	52704 .68	54337 .96	53346 .86
Yes	0	saddle point, Ψ	2333 .47	1260	.025	.930	.088	.096	52267 .57	54623 .27	53193 .79
Yes	1	saddle point, Ψ	2376 .84[b]	1284	.025	.928	.078	.079	52251 .70	54481 .77	53128 .53
Yes	2	saddle point, Ψ	2481 .46[c]	1332	.025	.925	.074	.076	52212 .85	54191 .64	52990 .88

Bayesian estimation (indicators treated as categorical)									
ind.-spec.					χ^2-diff. limit				
States	MI	Traceplots	PSR	PPP	lower	upper	Iterations	Thinning	Chains
No	0[a]	ok	1 .009	.000	187 .49	538 .74	60000	10	3
Yes	0	ok	1 .005	.020	6 .71	354 .29	60000	10	3
Yes	1	ok	1 .006	.015	21.02	367 .92	60000	10	3
Yes	2	ok	1 .004	.015	21.31	367 .03	60000	10	3

Note. AIC: Akaike information criterion; BIC: Bayesian information criterion; BIC adj.: sample size adjusted BIC; CFI: Comparative fit index; Ind.-spec. States: Model with indicator-specific states; MI: Level of measurement invariance: (0) no invariance restrictions on the loading or threshold parameters; (1) strong measurement invariance for the reference method (self-report) indicators, and invariant loadings of the method factors on the non-reference method indicators; (2) strong measurement invariance for all indicators; MLR: Maximum Likelihood robust estimator (Mplus); PPP: Posterior predictive p-value; Ψ: The latent variance-covariance matrix Ψ is not positive definite due to correlations between latent variables > 1; PSR: Potential scale reduction factor; RMSEA: Root mean square error of approximation; saddle point: the estimation algorithm has reached a saddle point; SRMR B: standardized root mean square residual on the between-level; SRMR W: standardized root mean square residual on the within-level.

[a] As the model with indicator-specific state factors fit the data better than the model with common latent state variables for MI level 0, further MI levels were not tested for the model with common latent state variables.

[b] χ^2-difference test (with MLR correction) with the less restrictive MI 0 model: $\chi^2(24) = 32.529$, $p = .114$.

[c] χ^2-difference test (with MLR correction) with the less restrictive MI 1 model: $\chi^2(48) = 132.94$, $p < .001$.

thereby setting the conditional method bias invariant over measurement occasions. Again, note that the residual variables $\varepsilon_{(r)tijkl}$ are set to 1 by definition. The following section presents the results of the model estimation.

8.3.1 Results

The LS-Com GRM with indicator-specific latent state variables and MI level 0 did fit the data slightly better than the respective LS-Com GRM with common state variables, as indicated by CFI, RMSEA, AIC, BIC and PPP, so that the model with indicator-specific states was chosen for further analyses. The specific values of the fit indices for the two models as estimated by MLR and Bayesian methods are provided in Table 8.2. Modification indices provided by the MLR estimation suggested that there were indicator-specific stabilities over time that could not be captured by the LS-Com GRM covariance structure. Model fit according to RMSEA and χ^2/df ratio indicated good fit, while SRMR values were acceptable and the CFI rather poor (as judged by the criteria given in Schermelleh-Engel et al. (2003), see section 8.2.3). PPP-values for the categorical-indicator model also indicated rather poor fit (PPP = 0.020), with the 95% confidence interval for the difference between the observed and replicated χ^2-values not including zero.

Despite the rather poor fit of the models, the results shall be shortly presented in the following in order to provide an illustration for how the coefficients in the LS-Com GRM could be interpreted. AIC and BIC values favored the more restrictive MI 2 model over the MI 0 and MI 1 models (see Table 8.2). Hence, the following results stem from the model with indicator-specific states and strong MI.

Table 8.3: Latent correlations between the latent state variables for the SWLS and SHS items in the LS-Com GRM

Scale		SWLS									SHS								
		S_{11}	S_{21}	S_{31}	S_{12}	S_{22}	S_{32}	S_{13}	S_{23}	S_{33}	H_{11}	H_{21}	H_{31}	H_{12}	H_{22}	H_{32}	H_{13}	H_{23}	H_{33}
SWLS	S_{11}	*3.167*	[.546, .725]	[.921, .973]	[.730, .857]	[.442, .652]	[.695, .831]	[.650, .802]	[.383, .608]	[.601, .761]	[.623, .795]	[.614, .772]	[.624, .777]	[.532, .721]	[.507, .690]	[.444, .641]	[.426, .632]	[.446, .642]	[.416, .618]
	S_{21}	.640	*2.472*	[.642, .793]	[.431, .638]	[.838, .937]	[.472, .666]	[.386, .607]	[.708, .859]	[.444, .645]	[.411, .639]	[.343, .564]	[.326, .547]	[.346, .579]	[.313, .538]	[.242, .475]	[.296, .529]	[.294, .520]	[.208, .445]
	S_{31}	.951	.722	*4.558*	[.752, .871]	[.564, .740]	[.773, .878]	[.650, .800]	[.473, .674]	[.663, .801]	[.742, .872]	[.713, .839]	[.714, .840]	[.655, .808]	[.633, .784]	[.565, .732]	[.542, .718]	[.552, .721]	[.528, .702]
	S_{12}	.798	.539	.816	*3.685*	[.569, .745]	[.933, .979]	[.739, .864]	[.458, .664]	[.738, .860]	[.526, .724]	[.485, .674]	[.546, .720]	[.758, .880]	[.651, .797]	[.629, .780]	[.540, .722]	[.525, .703]	[.567, .734]
	S_{22}	.552	.893	.657	.662	*2.538*	[.628, .782]	[.406, .621]	[.701, .856]	[.478, .673]	[.348, .584]	[.250, .484]	[.258, .492]	[.459, .665]	[.375, .588]	[.307, .532]	[.299, .533]	[.262, .496]	[.231, .466]
	S_{32}	.768	.574	.830	.960	.710	*5.195*	[.691, .828]	[.469, .667]	[.735, .852]	[.597, .774]	[.567, .731]	[.593, .754]	[.832, .924]	[.747, .861]	[.693, .825]	[.600, .762]	[.587, .747]	[.617, .770]
	S_{13}	.731	.501	.730	.807	.518	.764	*3.774*	[.673, .826]	[.926, .975]	[.531, .728]	[.493, .681]	[.462, .657]	[.609, .780]	[.529, .708]	[.501, .689]	[.694, .833]	[.645, .794]	[.585, .750]
	S_{23}	.500	.790	.579	.566	.785	.573	.756	*2.484*	[.730, .857]	[.367, .605]	[.257, .497]	[.231, .473]	[.391, .616]	[.300, .531]	[.270, .506]	[.502, .697]	[.422, .632]	[.334, .558]
	S_{33}	.685	.549	.737	.804	.581	.798	.953	.798	*5.469*	[.633, .799]	[.578, .741]	[.571, .735]	[.728, .858]	[.647, .793]	[.642, .789]	[.813, .909]	[.748, .865]	[.720, .842]
SHS	H_{11}	.715	.530	.813	.630	.471	.691	.635	.492	.721	*2.382*	[.867, .962]	[.843, .949]	[.748, .889]	[.783, .914]	[.735, .884]	[.771, .898]	[.733, .896]	[.742, .888]
	H_{21}	.698	.457	.780	.584	.370	.653	.591	.381	.664	.922	*3.155*	[.839, .937]	[.724, .864]	[.837, .927]	[.741, .872]	[.756, .882]	[.818, .918]	[.753, .881]
	H_{31}	.705	.441	.781	.637	.379	.678	.564	.356	.658	.902	.892	*3.191*	[.752, .883]	[.782, .903]	[.846, .932]	[.712, .852]	[.723, .862]	[.818, .916]
	H_{12}	.632	.467	.737	.824	.568	.883	.700	.509	.798	.825	.799	.823	*2.569*	[.908, .969]	[.880, .959]	[.809, .915]	[.785, .907]	[.844, .937]
	H_{22}	.603	.430	.713	.729	.486	.808	.623	.420	.724	.855	.886	.847	.943	*3.590*	[.884, .959]	[.817, .918]	[.852, .937]	[.845, .940]
	H_{32}	.547	.362	.653	.709	.424	.764	.599	.392	.720	.817	.811	.893	.925	.926	*3.492*	[.794, .909]	[.799, .910]	[.915, .971]
	H_{13}	.533	.416	.634	.636	.420	.685	.768	.605	.865	.840	.825	.787	.867	.872	.857	*3.397*	[.916, .976]	[.887, .959]
	H_{23}	.549	.411	.641	.619	.383	.671	.725	.532	.811	.822	.873	.797	.851	.899	.860	.952	*3.789*	[.879, .956]
	H_{33}	.521	.330	.619	.655	.351	.698	.673	.451	.785	.821	.822	.872	.895	.897	.947	.928	.922	*3.695*

Note. The lower diagonal contains posterior means, the upper diagonal posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonal. H_{il}: Happiness (SHS) latent state variable for indicator i on measurement occasion l; S_{il}: SWLS latent state variable for indicator i on measurement occasion l; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale.

Table 8.4: Latent correlations between the latent method variables for the SWLS and SHS items in the LS-Com GRM

		Common Method Variables						Method Variables					
Scale		SWLS			SHS			SWLS			SHS		
		CM_{S1}	CM_{S2}	CM_{S3}	CM_{H1}	CM_{H2}	CM_{H3}	M_{S1}	M_{S2}	M_{S3}	M_{H1}	M_{H2}	M_{H3}
	CM_{S1}	*0.463*	[.598, .881]	[.520, .850]	[.595, .874]	[.488, .857]	[.433, .829]	[.013, .437]	[.044, .462]	[-.005, .431]	[-.062, .347]	[-.145, .274]	[-.138, .283]
SWLS	CM_{S2}	.765	*0.409*	[.586, .883]	[.488, .847]	[.513, .869]	[.516, .868]	[-.038, .427]	[.081, .559]	[-.048, .461]	[-.043, .424]	[-.031, .454]	[-.111, .386]
	CM_{S3}	.710	.764	*0.477*	[.468, .842]	[.364, .819]	[.511, .875]	[-.111, .358]	[-.039, .447]	[-.195, .291]	[-.007, .441]	[-.009, .457]	[-.126 .343]
	CM_{H1}	.760	.700	.685	*0.378*	[.578, .875]	[.529, .860]	[-.108, .319]	[-.052, .388]	[-.114, .341]	[-.001, .396]	[-.077 .349]	[-.075 .350]
SHS	CM_{H2}	.707	.731	.631	.757	*0.343*	[.578, .888]	[-.112, .369]	[-.028, .469]	[-.105, .409]	[-.144, .335]	[-.192, .310]	[-.233, .265]
	CM_{H3}	.662	.730	.733	.727	.773	*0.345*	[-.146, .356]	[-.059, .463]	[-.248, .288]	[-.064, .411]	[-.093, .407]	[-.232, .277]
	M_{S1}	.232	.200	.126	.107	.132	.108	*1.078*	[.555, .766]	[.558, .772]	[.667, .833]	[.381, .632]	[.429, .675]
SWLS	M_{S2}	.259	.330	.209	.172	.228	.211	.668	*0.883*	[.716, .876]	[.482, .708]	[.587, .779]	[.505, .734]
	M_{S3}	.220	.214	.048	.116	.156	.019	.672	.803	*1.112*	[.403, .647]	[.392, .642]	[.609, .791]
	M_{H1}	.146	.198	.221	.202	.101	.180	.762	.602	.531	*1.259*	[.728, .881]	[.708, .868]
SHS	M_{H2}	.065	.220	.231	.139	.063	.162	.513	.690	.524	.812	*1.214*	[.744, .896]
	M_{H3}	.073	.142	.112	.140	.017	.022	.559	.626	.707	.796	.872	*1.358*

Note. The lower diagonal contains posterior means, the upper diagonal posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonal. CM_{Hl}: Common Method factor of SHS on measurement occasion l; CM_{Sl}: Common Method factor of SWLS on measurement occasion l; M_{Hl}: Method factor of SHS on measurement occasion l; M_{Sl}: Method factor of SWLS on measurement occasion l; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale.

Table 8.5: Latent correlations between the unique method variables for the SWLS and SHS items in the LS-Com GRM

Scale		SWLS			SHS		
		UM_{S1}	UM_{S2}	UM_{S3}	UM_{H1}	UM_{H2}	UM_{H3}
	UM_{S1}	*1.046*	[.715, .852]	[.651, .817]	[.514, .691]	[.350, .570]	[.326, .557]
SWLS	UM_{S2}	.788	*1.317*	[.746, .878]	[.432, .639]	[.517, .696]	[.431, .642]
	UM_{S3}	.739	.817	*1.353*	[.337, .572]	[.311, .545]	[.473, .668]
	UM_{H1}	.607	.540	.459	*0.676*	[.731, .876]	[.715, .873]
SHS	UM_{H2}	.464	.611	.432	.810	*0.796*	[.754, .892]
	UM_{H3}	.446	.542	.576	.800	.830	*0.860*

Note. The lower diagonal contains posterior means, the upper diagonal posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonal. SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; UM_{Hl}: Unique method factor of SHS on measurement occasion l; UM_{Sl}: Unique method factor of SWLS on measurement occasion l.

Posterior means and CIs for the correlations between the latent state variables of different measurement occasions as well as constructs are given in Table 8.3. The items of the SHS were observed to be rather homogeneous, with high correlations between the item-specific latent state variables on the same measurement occasion (.892 to .952). The items of the SWLS showed less homogeneity, with the latent states of Item 1 and Item 3 being highly correlated within measurement occasions (r = .951 - .960), while Item 2 showed less associations with the other two items (r = .640 - .798).

The indicator-specific latent state variables of the SWLS exhibited stabilities of .731 to .893, the stabilities of the indicator-specific SHS state variables ranged from .825 to .947. Construct stability corrected for indicator-specific effects, as indicated by correlations between the latent state variables of different indicators of the same construct over time, were slightly lower, albeit still high, ranging from .500 to .816 for the SWLS, and .787 to .897 for the SHS. Mean differences in satisfaction with life and happiness over time were found for SWLS item 1 (latent state mean of 0.307, CI: [0.125; 0.490], at T2 and 0.497, CI: [0.300; 0.698], at T3) and SWLS item 2 at T3 (latent state mean of 0.229, CI: [0.049;0.413]).

Correlations between SWLS and SHS item-specific latent states on the same measurement occasion ranged between .424 and .865. Correlations of SWLS and SHS latent states of different measurement occasions, indicating discriminant validity corrected for occasion-specific influences, were somewhat lower, ranging from .330 to .798.

Posterior means and CIs for the correlations between the latent common method factors and method factors of different measurement occasions as well as constructs are given in Table 8.4, and those for the unique method factors are given in Table 8.5. Stability of rater effects for the peer ratings were comparable for life satisfaction and subjective happiness: unique method factors exhibited stabilities of .788 and .817 between adjacent occasions for the SWLS and of .810 and .830 for the SHS; common method factor stabilities were .764 and .765 between adjacent occasions for the SWLS and .757 as well as .773 for the SHS. Method factor stabilities for the rater effects of the parents showed stabilities of .668 and .803 between adjacent measurement occasions for the SWLS and of .812 and .872 for the SHS. Rater effects also showed comparable stabilities between T1 and T3 (see Tables 8.4 and 8.5).

Generalizability of method effects across the two well-being aspects on the same measurement occasion was high for the unique method variables (.576 - .611), as well as the common method variables (.731 - .760) and method variables (.690 - .762). Generalizability of method effects across the two well-being aspects over measurement occasions, i.e., corrected for common occasion-specific effects, were still high, with correlations between .432 and .542 for the unique method, .631 and .730 for the common method, and .513 and .626 for the method factors.

Table 8.6: Consistency coefficients and method specificty coefficients for the SWLS and SHS items

	SWLS									SHS								
	item 1			item 2			item 3			item 1			item 2			item 3		
Coefficient	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Peer report																		
CON	.191	.194	.189	.192	.163	.156	.199	.195	.195	.151	.151	.182	.136	.138	.136	.169	.168	.167
	[.133, .255]	[.134, .259]	[.128, .255]	[.128, .264]	[.107, .228]	[.101, .222]	[.143, .260]	[.139, .258]	[.138, .260]	[.093, .216]	[.094, .216]	[.116, .254]	[.087, .191]	[.088, .195]	[.088, .192]	[.111, .235]	[.111, .232]	[.110, .232]
CMS	.248	.191	.212	.003	.002	.001	.198	.153	.168	.304	.255	.234	.183	.149	.141	.235	.195	.186
	[.144, .365]	[.107, .294]	[.122, .319]	[-.070, .091]	[-.053, .066]	[-.065, .071]	[.101, .308]	[.070, .256]	[.083, .271]	[.191, .429]	[.140, .389]	[.127, .358]	[.106, .274]	[.072, .245]	[.066, .235]	[.122, .357]	[.084, .324]	[.076, .320]
UMS	.561	.615	.600	.805	.836	.843	.603	.652	.637	.545	.594	.584	.681	.714	.723	.596	.636	.646
	[.449, .672]	[.513, .712]	[.496, .698]	[.694, .909]	[.747, .918]	[.750, .931]	[.493, .712]	[.550, .749]	[.535, .733]	[.430, .660]	[.465, .715]	[.463, .700]	[.584, .771]	[.610, .805]	[.620, .812]	[.484, .708]	[.512, .751]	[.516, .763]
Parent report																		
CON	.352	.434	.385	.301	.351	.296	.314	.389	.348	.266	.288	.323	.278	.312	.300	.318	.345	.333
	[.257, .452]	[.328, .542]	[.284, .490]	[.208, .401]	[.244, .462]	[.202, .399]	[.235, .398]	[.299, .482]	[.262, .438]	[.171, .372]	[.189, .397]	[.215, .437]	[.188, .378]	[.217, .415]	[.208, .401]	[.222, .420]	[.246, .450]	[.235, .438]
MS	.648	.566	.491	.587	.545	.473	.686	.611	.521	.734	.712	.677	.646	.616	.564	.682	.655	.600
	[.548, .743]	[.458, .672]	[.370, .639]	[.463, .726]	[.423, .685]	[.340, .638]	[.602, .765]	[.518, .701]	[.399, .666]	[.628, .829]	[.603, .811]	[.563, .785]	[.517, .793]	[.491, .758]	[.413, .751]	[.580, .778]	[.550, .754]	[.457, .771]
Latent State Means																		
	0 [a]	.161	.257	0 [a]	.046	.146	0 [a]	.042	.077	0 [a]	-.104	-.004	0 [a]	.012	-.012	0 [a]	.086	.084
	-	[.066, .256]	[.154, .362]	-	[-.057, .148]	[.032, .259]	-	[-.046, .129]	[-.019, .172]	-	[-.215, .004]	[-.102, .093]	-	[-.075, .100]	[-.101, .075]	-	[-.001, .174]	[-.005, .172]

Note. Credibility intervals are given in parentheses below the posterior mean values. CMS: Common Method Specificity Coefficient; CON: Consistency Coefficient; MS: Method Specificity Coefficient; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; T1: measurement occasion 1; T2: measurement occasion 2; T3: measurement occasion 3; UMS: Unique Method Specificity Coefficient.

[a] Latent State means at T1 were fixed to zero for identification reasons.

Correlations between common method factors and (parent) method factors were close to zero (with CIs including zero) for both the SWLS and SHS on most measurement occasions, with the exception of correlations for the SWLS at T1 and T2 of .232 and .330, respectively.
Estimates of consistency and method specificity coefficients along with their 95% CI for the items of the SWLS and SHS are presented in Table 8.6. Consistency coefficients for the SWLS peer report indicators range from .156 to .199. That is, depending on the item and measurement occasion, between 15.6% and 19.9% of the reliable variance of the peer reports can be explained by the self-reports on the SWLS. Consistency coefficients of the SHS peer report items are minimally smaller, where 13.6% to 18.2% of the variance of the latent response variable is shared with the self-report latent state variables. The square root of the consistency coefficient can be interpreted as the latent correlation between the self- and the peer report. That is, in the present sample and model, SWLS self- and peer reports correlate to .39 - .45, SHS self- and peer reports correlate to .37 - .43. The largest amount of the variance in the peer reports is variance that goes back to the individual views of the peers, not shared with either the self-report or the other peers. This is the case for the SWLS and the SHS, as indicated by unique method specificity coefficients of .561 - .843 (SWLS) and .545 - .723 (SHS). In contrast, a common view of the peers that is not shared with the self-report accounts for only 0.1% - 24.8% of the interindividual differences in the SWLS peer reports and for 14.1% - 30.4% of the variance in the SHS peer reports.
In contrast to the peer reports, the parent reports show higher correlations with the self-reports for both SWLS and SHS. For the SWLS parent report indicators between 29.6% and 43.4% of the latent response variable's variance is shared with the self-report latent state variables, for the SHS between 26.6% and 34.5%.

8.3.2 Discussion

As the LS-Com GRM did not fit the data well, the results presented above should be interpreted with caution. Nevertheless, they give some hints with respect to the amount and generalizability of different rater effects. Consistency and method specificity coefficients of the SWLS and SHS items indicate that it is not enough to collect only self-, parent or peer reports, but that several different methods should be considered. Observed rater effects seem to generalize over the two components, that is, peers or parents (under-) overestimating the life satisfaction of a target tend to also (under-) overestimate the happiness of the respective target. This is not only the case for individual raters (unique method factor) but also for the common view of the peers on the target (common method factor). Furthermore, these rater tendencies of over- or underestimation appear to be quite stable over a time period of three to six months. Also, both life satisfaction and subjective happiness themselves were found to be quite stable, however, still indicating variability in both components over time. However, modification indices indicated the existence of indicator-specific stabilities that are not well recovered by the LS-Com GRM covariance structure. These might be better captured by the LST-Com GRM.
Furthermore, the LS-Com GRM does not allow for the differentiation between state variability and trait change. In order to make this distinction, other models such as latent state-trait-models are needed (Eid & Kutscher, 2014). A latent state-trait version of the model, the LST-Com GRM, will be applied to the data in the following section.

8.4 LST-Com GRM of Life Satisfaction and Subjective Happiness

In this section the LST-Com GRM is applied to the SWLS and SHS self-, parent and peer report measures. The specific LST-Com GRM used corresponds to the model depicted in Figure 8.1, with the only difference that the model in this application includes three measurement occasions (instead of two, as depicted in the Figure). Due to the results of the application of the LS-Com GRM to these data (i.e., indicator-specific effects and stabilities), the LST-Com GRM was specified with indicator-specific latent trait variables ξ_{tij1} and ξ^{UM}_{rtij2}, ξ^{CM}_{tij2}, and ξ^{M}_{tij3}.

Different levels of measurement invariance were compared for the chosen model by a stepwise approach. The levels tested were the following three levels: (0) factorial invariance, i.e., no invariance restrictions on the loading or threshold parameters; (1) invariance of the loading parameters of the latent state residual (method) factors on the non-reference method indicators, i.e., invariant $\lambda_{\zeta ijkl}$, $\lambda^{CM}_{\zeta ijkl}$, $\lambda^{M}_{\zeta ijkl}$, and $\lambda^{UM}_{\zeta ijkl}$ over time; (2) strong measurement invariance, i.e., all factor loadings and threshold parameters restricted to be equal across measurement occasions for the same indicator i, construct j and method k. This includes the regression parameters of the regression of the non-reference method on the reference-method trait variables, thereby setting the conditional method bias invariant over measurement occasions. Again, note that the residual variables $\varepsilon_{(r)tijkl}$ are set to 1 by definition. The following section presents the results of the model estimation.

8.4.1 Results

Model fit for the LST-Com GRM with MI levels 0, 1 and 2 as estimated by MLR and Bayesian methods are given in Table 8.7. The model with MI 0 (no measurement invariance restrictions) did either not converge (MLR) or showed problematic traceplots indicating poor convergence for some parameters (Bayesian estimation). So did the traceplots of the model with MI 1. Traceplots for the MI 0 and MI 1 model did not show patterns suggesting that further increasing the number of iterations would improve convergence. AIC and BIC values favored the more restrictive model with MI 2 (see Table 8.7). The LST-Com GRM with MI 2 converged well, as indicated by the inspection of traceplots and a PSR of 1.008. Hence, the following results stem from the Bayesian estimation of the model with strong MI (MI level 2). The model fits the data reasonably well, with a PPP-value of .137 and a 95% confidence interval for the difference between the observed and replicated χ^2-values including zero, [-74.08, 265.87]. Estimates of reliability, ICCs, and trait- as well as occasion-specificities for the items of the SWLS and SHS are presented in Table 8.8. Reliabilities of the items are satisfactory, ranging from .721 to .889 for the SWLS and .695 to .817 for the SHS items, with no marked differences between the self-, parent or peer reports. True (i.e., latent) ICC estimates show that the need for modeling the multilevel structure is clearly given.

Time consistency coefficients are high, ranging from .687 to .877 for the SWLS and from .767 to .927 for the SHS. That is, between 68.7% and 92.7% of the reliable variance of the items goes back to stable influences over time, with slightly lower occasion-specificity of the SHS than the SWLS items. No systematic differences in the time consistencies between self-, parent and peer reports can be observed. Figures 8.2 and 8.4 display category characteristic curves for the SWLS and SHS self-report, parent report and peer report items. Operation characteristic curves are displayed in Figures 8.3 and 8.5 for the SWLS and SHS, respectively.

For the SWLS items, the probability distributions of the different categories in dependency of the latent response variable are very similar for the self-reports and peer reports. That is, given a value on the latent response variable π_{rtijkl} or π_{tijkl}, the answer scale is not used in different ways by the targets and the peers. Parents, in contrast, show a general tendency to answering in higher categories

Table 8.7: Model fit results for the LST-Com GRM of life satisfaction and subjective happiness

MLR estimator (indicators treated as continuous)										
MI	remarks	χ^2	df	RMSEA	CFI	SRMR W	SRMR B	AIC	BIC	BIC adj.
0	no convergence	-	-	-	-	-	-	-	-	-
1	-	1786.883	1347	.015	.971	.033	.061	51621.50	53521.77	52368.66
2	-	1968.256[a]	1455	.016	.966	.040	.061	51601.21	52936.11	52126.07

Bayesian estimation (indicators treated as categorical)								
				χ^2-diff. limit				
MI	Traceplots	PSR	PPP	lower	upper	Iterations	Thinning	Chains
0	problematic	1.042	.265	-115.95	222.28	100000	10	3
1	problematic	1.023	.240	- 106.50	232.35	100000	10	3
2	good	1.008	.137	-74.08	265.87	80000	10	3

Note. AIC: Akaike information criterion; BIC: Bayesian information criterion; BIC adj.: sample size adjusted BIC; CFI: Comparative fit index; MI: Level of measurement invariance: (0) no measurement invariance restrictions on loading or threshold parameters; (1) invariance of the loading parameters of the latent state residual (method) factors on the non-reference method indicators, i.e., invariant $\lambda_{\zeta ijkl}$, $\lambda^{CM}_{\zeta ijkl}$, $\lambda^{M}_{\zeta ijkl}$, and $\lambda^{UM}_{\zeta ijkl}$ over time; (2) strong measurement invariance, i.e., all factor loadings and threshold parameters restricted to be equal across measurement occasions; MLR: Maximum Likelihood robust estimator (Mplus); PPP: Posterior predictive p-value; PSR: Potential scale reduction factor; RMSEA: Root mean square error of approximation; SRMR B: standardized root mean square residual on the between-level; SRMR W: standardized root mean square residual on the within-level.

[a] χ^2-difference test (with MLR correction) with the less restrictive MI 1 model: $\chi^2(108) = 174.02$, $p < .001$.

as compared to the targets (self-reports) or peers on the SWLS. SWLS item 2 appears to be slightly easier than Items 1 and 3, at least for Category 5.

For the SHS items, the probability distributions of the different categories in dependency of the latent response variable are similar for the self-reports, parent reports and peer reports. Only SHS item 2 appears to be minimally easier for the parents than for the peers or the targets (self-report). Category characteristic curves for SHS items 2 and 3 are comparable, while item 1 shows a different pattern. Item 1 seems to be the easiest of the SHS items, showing higher probabilities for categories 4 and 5, while category 2 is avoided.

Posterior means and CIs for the correlations between the latent trait variables, the latent state residual variables and the latent trait and state residual method variables in the LST-Com GRM are given in Tables 8.9, 8.10 and 8.11. In accordance with the correlations found for the latent state factors in the LS-Com GRM, the items of the SHS were observed to be rather homogeneous, with high correlations between the item-specific latent trait variables (.916 - .935; see Table 8.9). Again, the items of the SWLS showed less homogeneity, with the latent traits of Item 1 and Item 3 being highly correlated (.945), while Item 2 showed less associations with the other two items (r = .654 and .709).

Correlations between SWLS and SHS item-specific latent traits ranged between .439 and .860, with the lowest correlations found for SWLS item 2. That is, satisfaction with life as a trait shows strong associations with self-reported trait happiness. Note that, in contrast to the LS-Com GRM, these correlations refer to the stable part of the constructs and are free of measurement-occasion specific influences. However, also on the occasion-specific level self-reported satisfaction with life and subjective happiness are positively associated, as indicated by state residual correlations of .599 to .739 (see Table 8.10).

Correlations between the method trait variables of the SWLS and SHS items reported in Table 8.11 show that parents who consistently overestimate their child's satisfaction with life tend to also consistently overestimate their happiness, with slightly greater associations between the SHS with SWLS items 1 and 3 (.578 - .817) than with item 2 (.350 - .496). Similarly, if the stable common view of the peers on the targets' satisfaction with life is higher than is to be expected by the self-reports, they also tend to have a too positive common view on the targets' happiness (correlations between *CMT*

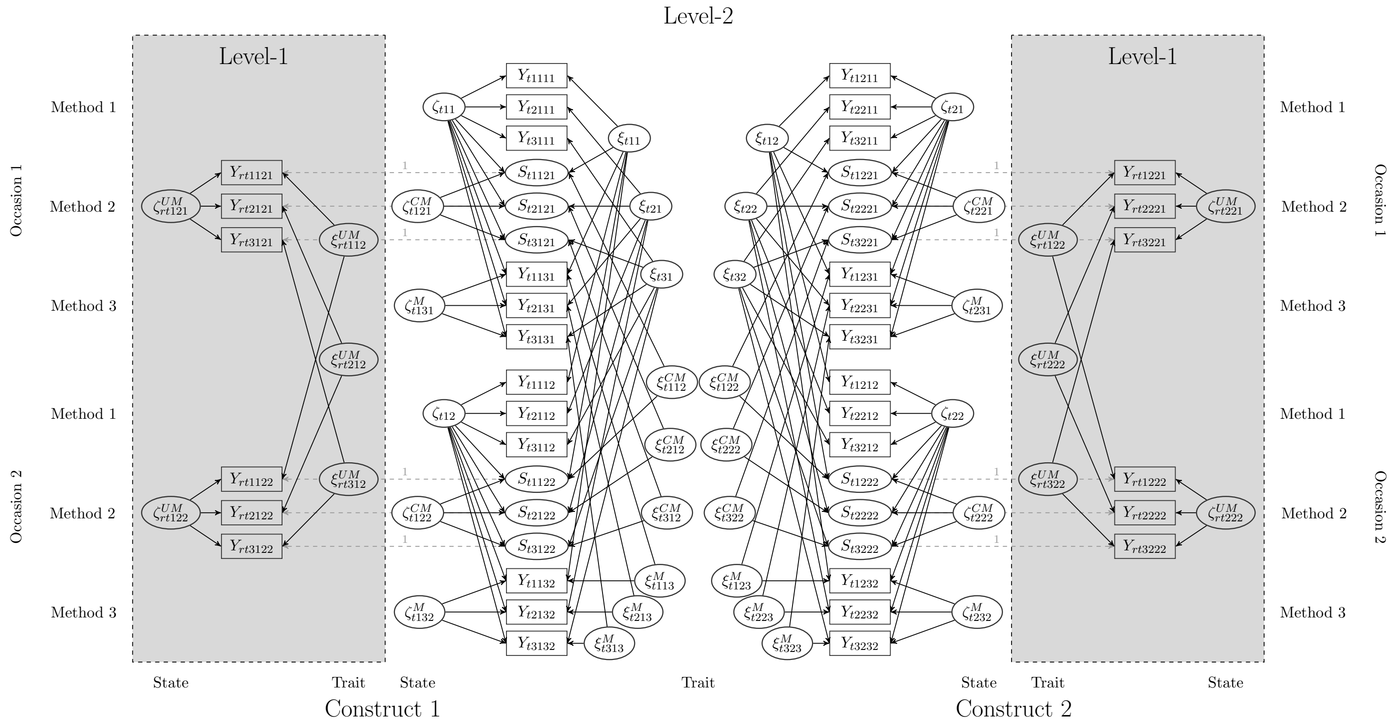

Figure 8.1: Path diagram of the Latent-State-Trait-Com graded response model with indicator-specific latent trait (ξ_{tij}) and latent method trait variables (ξ^{UM}_{rtij2}, ξ^{CM}_{tij2}, ξ^{M}_{tijk}), and common latent state residual variables ζ_{tjl}, two structurally different methods and one set of interchangeable methods on two measurement occasions for two constructs. Note that the model applied to the SWLS and SHS data included three measurement occasions (instead of two, as depicted here). Method 1 is selected as reference method. For the sake of clarity, correlations between latent variables and loading parameters are not displayed. Correlations that are not permissible in the depicted LST-Com GRM are all correlations between any trait (method) variable ξ and any state residual (method) variable ζ, correlations between the latent trait and the latent trait (common) method variables of the same construct j and indicator i, correlations between the latent state residual and the latent state residual (common) method variables of the same construct j and measurement occasion l, as well as correlations between any level-1 and any level-2 latent variable. Additionally restricted to zero were: any correlations of latent state residual (method) variables across measurement occasions; correlations between latent state residual variables and latent state residual method variables of different constructs; any correlations between latent trait variables and latent trait (common) method variables (across constructs). CM: common method; M: method; S: state variable; UM: unique method; ξ: latent trait variable; Y_{rtijkl}: rating of rater r for target t of the i-th item of trait j and method k on measurement occasion l; ζ: latent state residual variable.

variables of .511 - .784).

On the occasion-specific level, method correlations deviating from zero are found only for the parents' ratings of life satisfaction and happiness. Parents that have a momentary positively biased view on their child's satisfaction with life tend to also overestimate their happiness at that moment in time. Hence, generalizability of stable method effects across the two well-being aspects is high for the peers as well as the targets, while generalizability of transient method effects across life satisfaction and happiness is only possible for the parent ratings. However, generalizability of method effects for the peers is high on the individual rater level (see Table 8.10). That is, the specific view of a peer rater, not shared with the other peers, is highly generalizable over items within (.663 - .864) as well as between the two well-being aspects (.321 - .668) on a stable level, as well as between the two well-being aspects on a momentary level (.488 - .571).

The peers' stable common view on the target's satisfaction with life shares some similarities with the parents' stable view (correlations between .324 and .515), corrected for the self-reports. That is, targets whose satisfaction of life is overestimated (underestimated) by their parents, tend to be also overestimated (underestimated) in their satisfaction with life by their peers (common view of the peers). In contrast, this kind of association is not present for most of the happiness items (MT and

Table 8.8: Latent variance coefficients for the SWLS and SHS items in the LST-Com GRM: Reliabilities, ICCs, time consistency and occasion-specificity coefficients

					SWLS										SHS			
		item 1			item 2			item 3			item 1			item 2			item 3	
Report	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Reliabilities																		
Self	.784	.781	.802	.724	.721	.742	.868	.866	.880	.750	.739	.759	.798	.792	.804	.779	.773	.783
	[.737, .827]	[.735, .823]	[.757, .843]	[.671, .774]	[.667, .771]	[.688, .790]	[.827, .907]	[.825, .906]	[.842, .916]	[.701, .797]	[.691, .785]	[.712, .804]	[.756, .838]	[.749, .832]	[.762, .843]	[.736, .818]	[.731, .813]	[.742, .821]
Parent	.775	.748	.764	.844	.824	.835	.889	.869	.881	.749	.745	.752	.789	.784	.792	.729	.725	.731
	[.719, .825]	[.688, .802]	[.707, .815]	[.790, .891]	[.765, .876]	[.780, .884]	[.843, .930]	[.820, .915]	[.834, .925]	[.691, .803]	[.685, .799]	[.694, .806]	[.725, .848]	[.716, .848]	[.728, .851]	[.671, .783]	[.669, .778]	[.674, .784]
Peer	.779	.763	.779	.774	.772	.781	.793	.791	.804	.698	.695	.705	.815	.815	.817	.700	.700	.703
	[.729, .829]	[.714, .813]	[.730, .829]	[.729, .817]	[.725, .817]	[.735, .824]	[.740, .853]	[.740, .851]	[.753, .861]	[.647, .748]	[.643, .746]	[.653, .757]	[.751, .884]	[.753, .882]	[.757, .883]	[.653, .747]	[.652, .748]	[.655, .752]
ICC																		
Peer	.423	.357	.362	.323	.303	.299	.379	.359	.344	.407	.395	.416	.323	.321	.319	.354	.351	.352
	[.320, .525]	[.259, .462]	[.260, .470]	[.224, .427]	[.208, .407]	[.205, .402]	[.272, .482]	[.258, .462]	[.245, .445]	[.299, .512]	[.287, .503]	[.307, .523]	[.204, .433]	[.202, .432]	[.202, .428]	[.248, .460]	[.245, .458]	[.246, .460]
Time consistencies																		
Self	.766	.781	.687	.812	.824	.743	.766	.781	.687	.805	.853	.767	.863	.899	.834	.901	.927	.879
	[.686, .837]	[.707, .847]	[.602, .767]	[.737, .875]	[.751, .886]	[.656, .821]	[.691, .833]	[.700, .852]	[.611, .758]	[.723, .875]	[.780, .914]	[.678, .847]	[.802, .915]	[.843, .943]	[.762, .896]	[.848, .843]	[.881, .962]	[.822, .926]
Parent	.747	.867	.795	.759	.877	.809	.689	.827	.739	.821	.842	.808	.788	.812	.773	.884	.900	.875
	[.653, .831]	[.790, .927]	[.710, .867]	[.668, .842]	[.800, .936]	[.724, .881]	[.598, .775]	[.743, .898]	[.644, .826]	[735., .892]	[.752, .915]	[.708, .888]	[.692, .872]	[704., .905]	[.672, .862]	[.806, .944]	[.833, .951]	[.797, .935]
Peer	.719	.786	.718	.852	.864	.822	.776	.782	.724	.795	.806	.767	.795	.794	.782	.869	.869	.857
	[.639, .795]	[.710, .851]	[.632, .797]	[.791, .904]	[.795, .919]	[.746, .886]	[.683, .855]	[.693, .859]	[.634, .808]	[.735, .892]	[.730, .868]	[.683, .840]	[.696, .879]	[.704, .873]	[.693, .861]	[.803, .929]	[.797, .925]	[.780, .919]
Occasion specificities																		
Self	.234	.219	.313	.188	.176	.257	.234	.219	.313	.195	.147	.233	.137	.101	.166	.099	.073	.121
	[.163, .314]	[.153, .293]	[.233, .398]	[.125, .263]	[.114, .249]	[.179, .344]	[.167, .309]	[.148, .300]	[.242, .389]	[.125, .277]	[.086, .220]	[.153, .322]	[.085, .198]	[.057, .157]	[.104, .238]	[.057, .152]	[.038, .119]	[.074, .178]
Parent	.253	.133	.205	.241	.123	.191	.311	.173	.261	.179	.158	.192	.212	.188	.227	.116	.100	.125
	[.169, .247]	[.073, .210]	[.133, .290]	[.158, .332]	[.064, .200]	[.119, .276]	[.225, .402]	[.102, .257]	[.174, .356]	[.108, .265]	[.085, .248]	[.112, .292]	[.128, .308]	[.095, .296]	[.138, .328]	[.056, .193]	[.049, .167]	[.065, .203]
Peer	.281	.214	.282	.148	.136	.178	.224	.218	.276	.205	.194	.233	.205	.206	.218	.131	.131	.143
	[.205, .364]	[.149, .290]	[.203, .368]	[.096, .209]	[.081, .205]	[.114, .254]	[.145, .317]	[.141, .307]	[.192, .366]	[.142, .278]	[.132, .270]	[.160, .317]	[.121, .304]	[.127, .296]	[.139, .306]	[.077, .197]	[.075, .203]	[.081, .265]

Note. Credibility Intervals are given in parentheses below the posterior mean values. ICC: Intra-class-correlation coefficient; Report: Method, i.e., self-, parent or peer report; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; Tl: measurement occasion *l*. Time consistency and occasion-specificity coefficients were calculated by the formulas given in Table 4.1, reliabilities and ICCs as given by the formulas in Table 2.2.

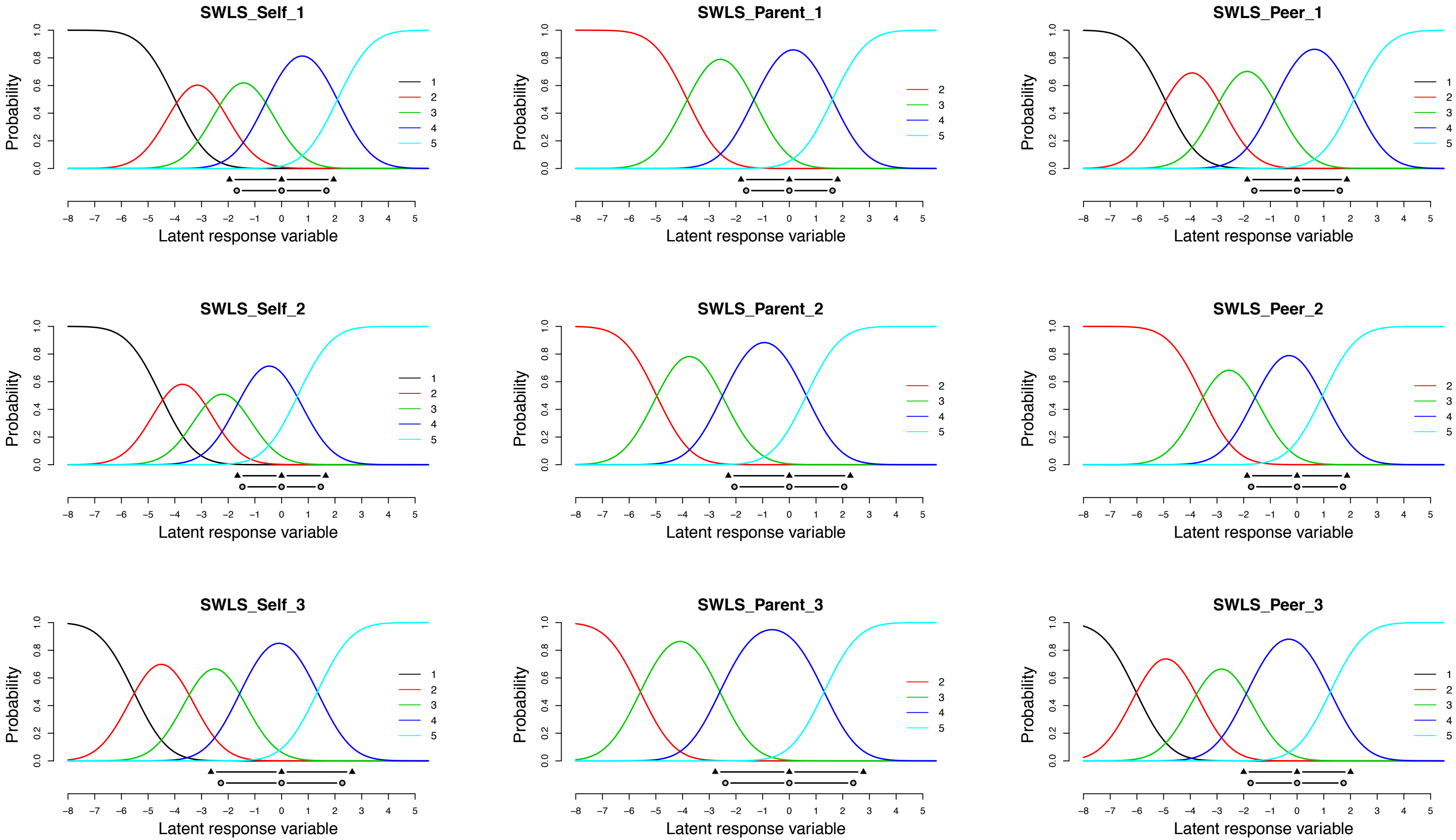

Figure 8.2: Category characteristic curves for the SWLS self-report, parent report and peer report items in the LST-COM GRM. Lines below the curves indicate the variability of the latent response variable π and of the trait components of π: Upper line with triangles: mean variability of π, averaged over measurement occasions; Lower line with circles: mean variability of the trait component of π (i.e., variability of π that goes back to trait components); middle triangle / circle: mean; lower / upper triangle / circle: mean +- 1 SD; Parent: parent report items; Peer; peer report items; Self: self-report items; SWLS: Satisfaction with Life Scale. Note that the lowest category of SWLS item 2 was not used by the peers at T3 and the lowest category of all SWLS items was not used by the parents on at least one occasion of measurement.

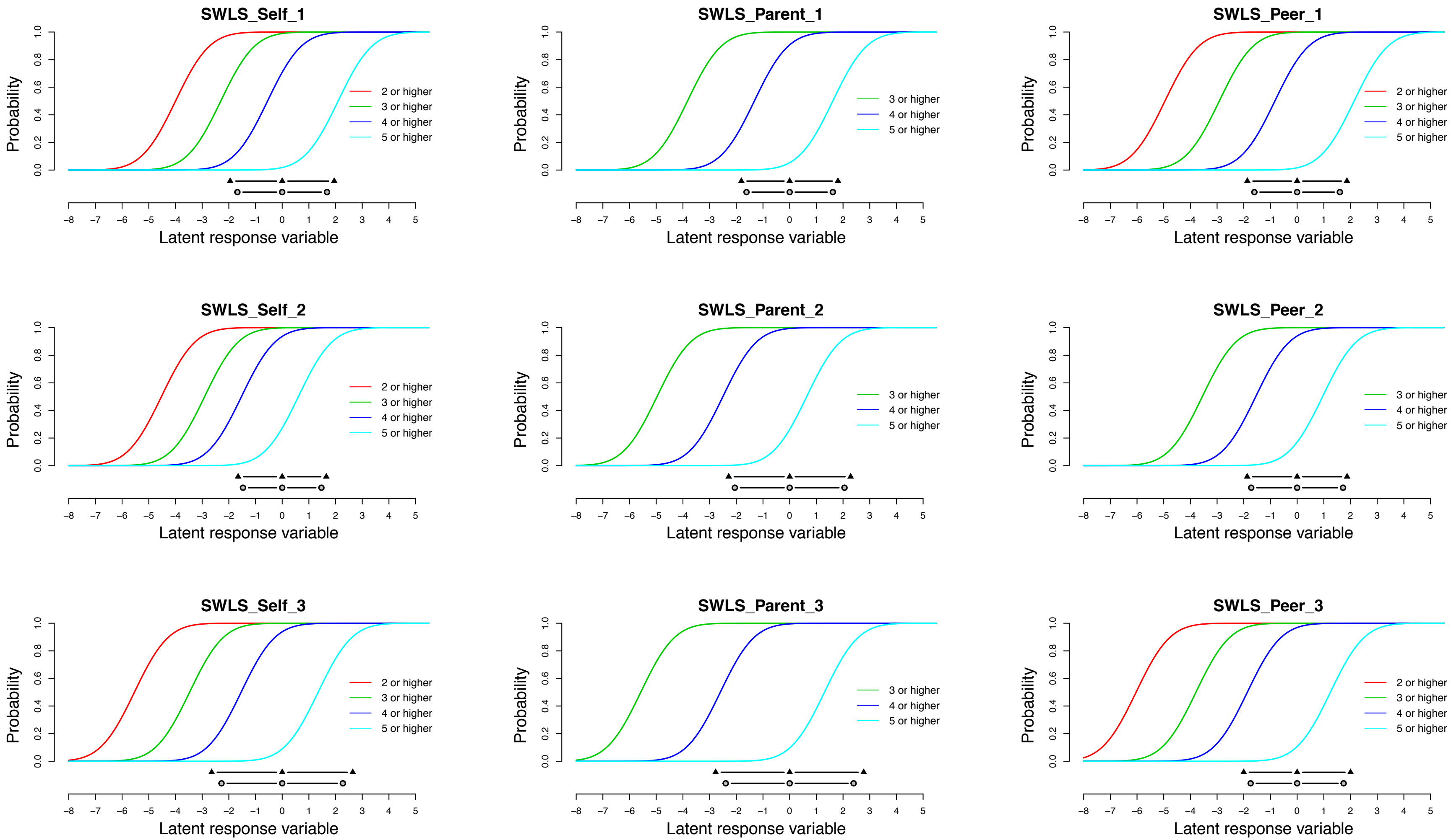

Figure 8.3: Operation characteristic curves for the SWLS self-report, parent report and peer report items in the LST-COM GRM. Lines below the curves indicate the variability of the latent response variable π and of the trait components of π: Upper line with triangles: mean variability of π, averaged over measurement occasions; Lower line with circles: mean variability of the trait component of π (i.e., variability of π that goes back to trait components); middle triangle / circle: mean; lower / upper triangle / circle: mean +- 1 SD; Parent: parent report items; Peer; peer report items; Self: self-report items; SWLS: Satisfaction with Life Scale. Note that the lowest category of SWLS item 2 was not used by the peers at T3 and the lowest category of all SWLS items was not used by the parents on at least one occasion of measurement.

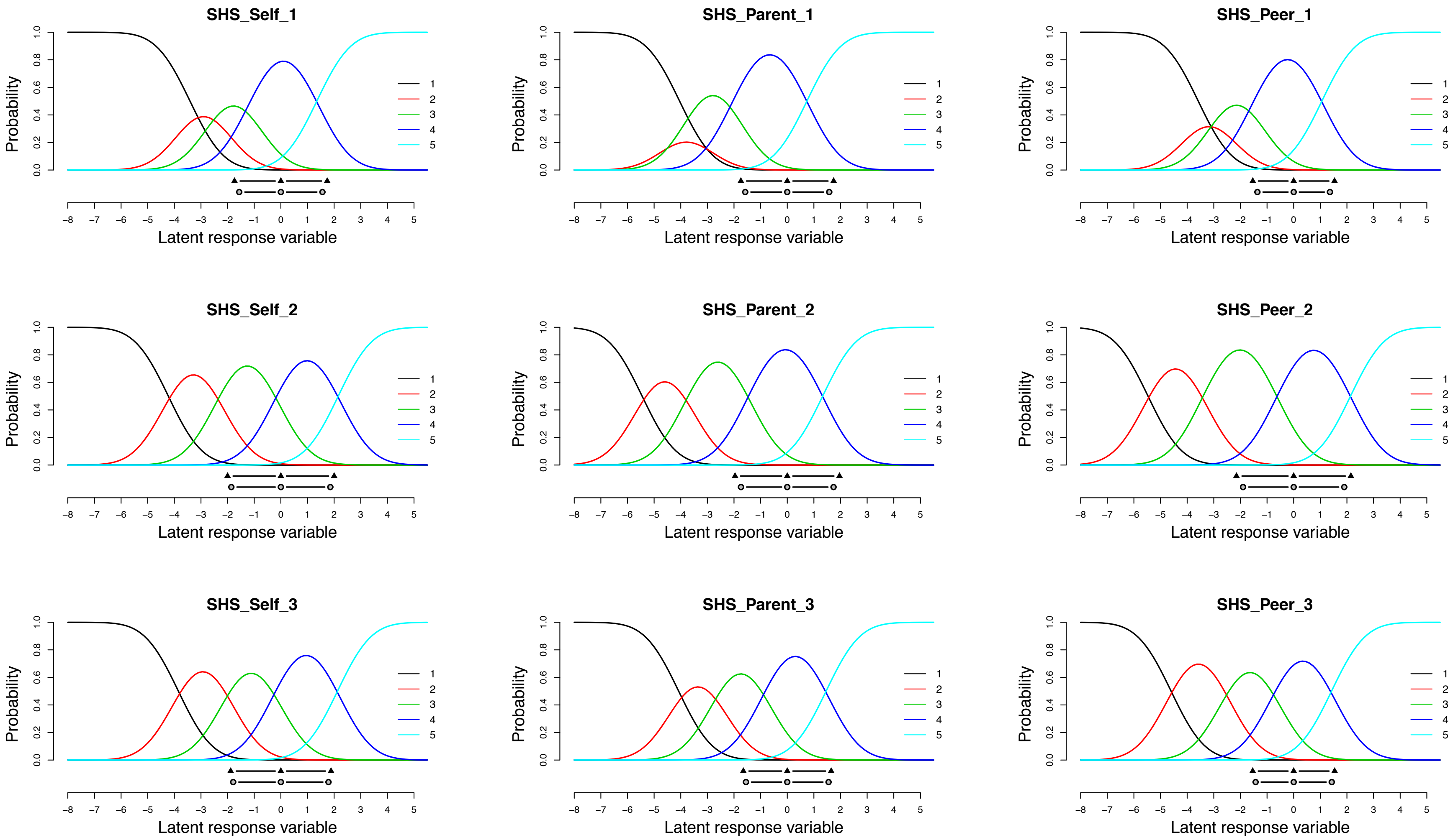

Figure 8.4: Category characteristic curves for the SHS self-report, parent report and peer report items in the LST-COM GRM. Lines below the curves indicate the variability of the latent response variable π and of the trait components of π: Upper line with triangles: mean variability of π, averaged over measurement occasions; Lower line with circles: mean variability of the trait component of π (i.e., variability of π that goes back to trait components); middle triangle / circle: mean; lower / upper triangle / circle: mean +- 1 SD; Parent: parent report items; Peer: peer report items; Self: self-report item;. SHS: Subjective Happiness Scale.

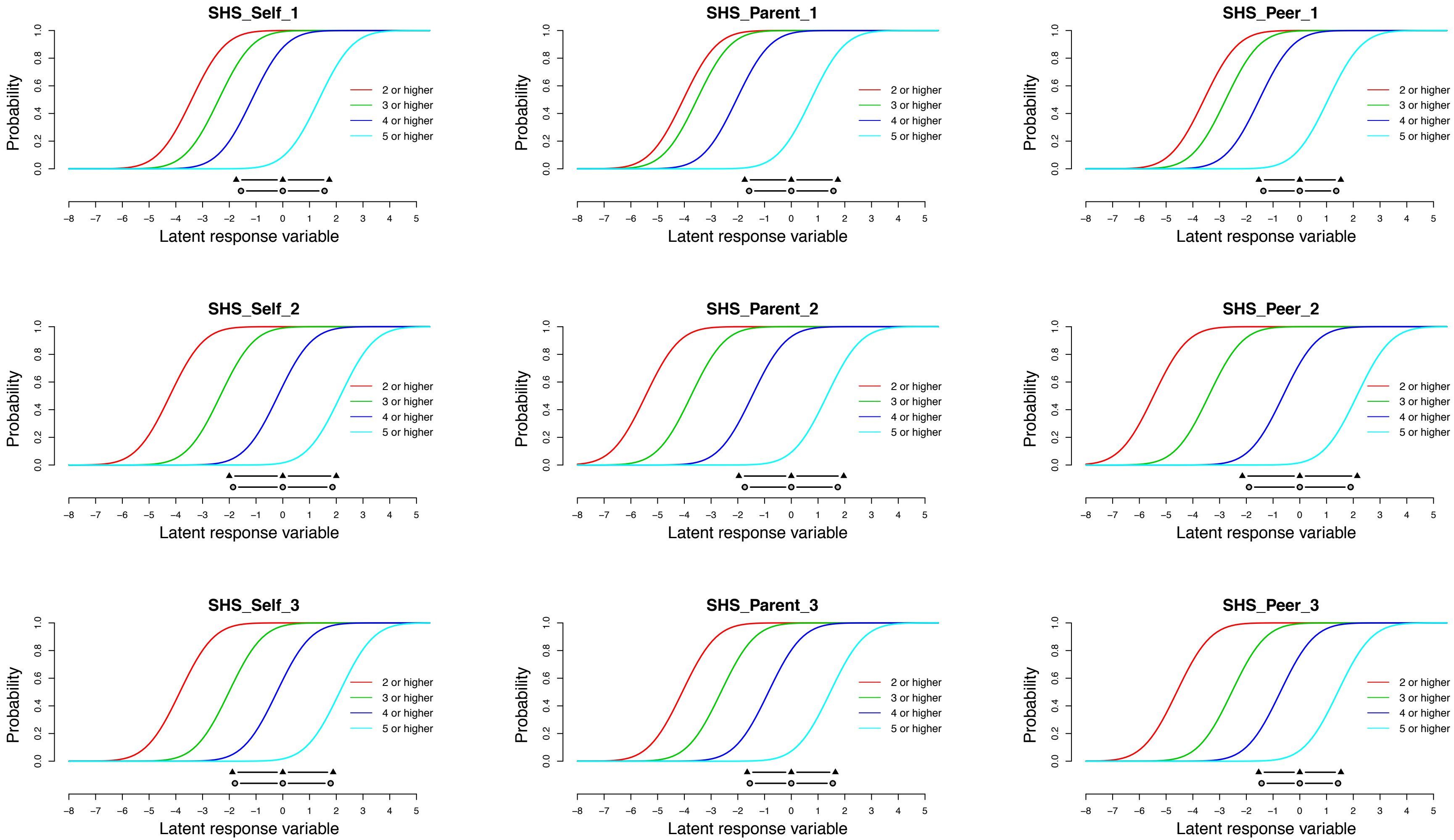

Figure 8.5: Operation characteristic curves for the SHS self-report, parent report and peer report items in the LST-COM GRM. Lines below the curves indicate the variability of the latent response variable π and of the trait components of π: Upper line with triangles: mean variability of π, averaged over measurement occasions; Lower line with circles: mean variability of the trait component of π (i.e., variability of π that goes back to trait components); middle triangle / circle: mean; lower / upper triangle / circle: mean +- 1 SD; Parent: parent report items; Peer: peer report items; Self: self-report items; SHS: Subjective Happiness Scale.

Table 8.9: Latent correlations between the SWLS and SHS trait variables in the LST-Com GRM

Scale		SWLS			SHS		
		T_{S1}	T_{S2}	T_{S3}	T_{H1}	T_{H2}	T_{H3}
SWLS	T_{S1}	*2.819*	[.574, .725]	[.920, .966]	[.706, .827]	[.654, .779]	[.632, .764]
	T_{S2}	.654	*2.185*	[.640, .769]	[.450, .629]	[.376, .560]	[.341, .530]
	T_{S3}	.945	.709	*5.170*	[.814, .900]	[.751, .849]	[.737, .839]
SHS	T_{H1}	.711	.544	.860	*2.441*	[.903, .961]	[.884, .944]
	T_{H2}	.720	.472	.803	.935	*3.457*	[.884, .944]
	T_{H3}	.702	.439	.791	.927	. 916	*3.200*

Note. The lower diagonal contains posterior means, the upper diagonal posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonal. SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; T_{Hi}: Latent trait variable for SHS item *i*; T_{Si}: Latent trait variable for SWLS item *i*.

Table 8.10: Latent correlations for the SWLS and SHS state residual, unique method state residual and unique method trait variables in the LST-Com GRM

Latent state residual variables							
Scale		SWLS			SHS		
		S_{S1}	S_{S2}	S_{S3}	S_{H1}	S_{H2}	S_{H3}
SWLS	S_{S1}	*0.867*	-	-	[.403, .768]	-	-
	S_{S2}	-	*0.791*	-	-	[.560, .872]	-
	S_{S3}	-	-	*1.295*	-	-	[.522, .814]
SHS	S_{H1}	.599	-	-	*0.598*	-	-
	S_{H2}	-	.739	-	-	*0.422*	-
	S_{H3}	-	-	.678	-	-	*0.750*
Unique method state residual variables							
Scale		SWLS			SHS		
		UMS_{S1}	UMS_{S2}	UMS_{S3}	UMS_{H1}	UMS_{H2}	UMS_{H3}
SWLS	UMS_{S1}	*0.487*	-	-	[.346, .762]	-	-
	UMS_{S2}	-	*0.530*	-	-	[.271, .708]	-
	UMS_{S3}	-	-	*0.715*	-	-	[.248, .701]
SHS	UMS_{H1}	.571	-	-	*0.298*	-	-
	UMS_{H2}	-	.503	-	-	*0.307*	-
	UMS_{H3}	-	-	.488	-	-	*0.328*
Latent trait variables							
Scale		SWLS			SHS		
		UMT_{S1}	UMT_{S2}	UMT_{S3}	UMT_{H1}	UMT_{H2}	UMT_{H3}
SWLS	UMT_{S1}	*1.578*	[.550, .767]	[.787, .925]	[.303, .594]	[.384, .645]	[.448, .692]
	UMT_{S2}	.663	*1.978*	[.642, .826]	[.171, .468]	[.179, .466]	[.178, .458]
	UMT_{S3}	.864	.740	*1.697*	[.482, .729]	[.485, .723]	[.554, .768]
SHS	UMT_{H1}	.454	.323	.613	*1.088*	[.688, .883]	[.695, .883]
	UMT_{H2}	.521	.326	.613	.793	*2.180*	[.771, .918]
	UMT_{H3}	.576	.321	.668	.796	.851	*1.241*

Note. The lower diagonals contain posterior means, the upper diagonals posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonals. SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; S_{Hl}: SHS latent state residual variable on measurement occasion *l*; S_{Sl}: SWLS latent state residual variable on measurement occasion *l*; UMS_{Hl}: SHS unique method state residual variable on measurement occasion *l*; UMS_{Sl}: SWLS unique method state residual variable on measurement occasion *l*; UMT_{Hi}: SHS unique method trait variable of item *i*; UMT_{Si}: SWLS unique method trait variable of item *i*.

Table 8.11: Latent correlations between the SWLS and SHS common method and method state residual and trait variables in the LST-Com GRM

Scale													
Method trait variables													
Scale			SWLS			SHS			SWLS			SHS	
		CMT_{S1}	CMT_{S2}	CMT_{S3}	CMT_{H1}	CMT_{H2}	CMT_{H3}	MT_{S1}	MT_{S2}	MT_{S3}	MT_{H1}	MT_{H2}	MT_{H3}
SWLS	CMT_{S1}	*0.449*	[.207, .838]	[.558, .914]	[.305, .855]	[.242, .859]	[.120, .804]	[.169, .763]	[.123, .725]	[.168, .754]	[.043, .642]	[.094, .657]	[.014, .604]
	CMT_{S2}	.603	*0.555*	[.370, .867]	[.265, .844]	[.231, .846]	[.077, .781]	[.036, .607]	[.178, .727]	[.055, .614]	[-.033, .516]	[.080, .596]	[-.095, .453]
	CMT_{S3}	.795	.687	*0.609*	[.554, .905]	[.494, .905]	[.323, .850]	[.245, .778]	[.130, .691]	[.196, .750]	[.042, .601]	[.239, .734]	[.120, .657]
SHS	CMT_{H1}	.660	.629	.784	*0.427*	[.488, .893]	[.324, .840]	[.106, .638]	[.115, .657]	[.058, .594]	[-.125, .413]	[.137, .619]	[-.051, .465]
	CMT_{H2}	.648	.624	.771	.757	*0.819*	[.535, .907]	[.087, .654]	[-.025, .546]	[-.105, .460]	[-.176, .372]	[-.016, .502]	[-.029, .501]
	CMT_{H3}	.553	.511	.663	.657	.788	*0.395*	[.085, .662]	[-.147, .445]	[-.121, .472]	[-.077, .504]	[.035, .578]	[.019, .573]
SWLS	MT_{S1}	.480	.324	.515	.375	.360	.377	*1.635*	[.492, .737]	[.680, .861]	[.480, .738]	[.444, .699]	[.460, .715]
	MT_{S2}	.428	.457	.402	.381	.243	.141	.621	*2.970*	[.674, .853]	[.354, .625]	[.278, .553]	[.200, .489]
	MT_{S3}	.465	.333	.464	.322	.161	.164	.779	.772	*3.245*	[.726, .893]	[.645, .838]	[.568, .791]
SHS	MT_{H1}	.345	.243	.315	.140	.094	.212	.616	.496	.817	*1.672*	[.684, .868]	[.713, .886]
	MT_{H2}	.379	.343	.490	.382	.237	.307	.578	.422	.749	.783	*2.090*	[.745, .902]
	MT_{H3}	.309	.174	.383	.205	.230	.296	.594	.350	.686	.807	.830	*1.602*
Method state residual variables													
Scale			SWLS			SHS			SWLS			SHS	
		CMS_{S1}	CMS_{S2}	CMS_{S3}	CMS_{H1}	CMS_{H2}	CMS_{H3}	MS_{S1}	MS_{S2}	MS_{S3}	MS_{H1}	MS_{H2}	MS_{H3}
SWLS	CMS_{S1}	*0.505*	-	-	[-.377, .645]	-	-	[-.056, .621]	-	-	[-.321, .480]	-	-
	CMS_{S2}	-	*0.153*	-	-	[-.432, .506]	-	-	[-.530, .493]	-	-	[-.468, .564]	-
	CMS_{S3}	-	-	*0.267*	-	-	[-.309, .656]	-	-	[-.647, .315]	-	-	[-.613, .374]
SHS	CMS_{H1}	.178	-	-	*0.173*	-	-	[-.234, .594]	-	-	[-.248, .605]	-	-
	CMS_{H2}	-	.046	-	-	*0.136*	-	-	[-.566, .411]	-	-	[-.547, .438]	-
	CMS_{H3}	-	-	.221	-	-	*0.229*	-	-	[-.810, -.155]	-	-	[-.788, -.105]
SWLS	MS_{S1}	.296	-	-	.206	-	-	*0.865*	-	-	[.548, .873]	-	-
	MS_{S2}	-	-.026	-	-	-.101	-	-	*0.379*	-	-	[.258, .779]	-
	MS_{S3}	-	-	-.201	-	-	-.549	-	-	*0.636*	-	-	[.477, .848]
SHS	MS_{H1}	.073	-	-	.206	-	-	.735	-	-	*0.539*	-	-
	MS_{H2}	-	.060	-	-	-.069	-	-	.552	-	-	*0.467*	-
	MS_{H3}	-	-	-.141	-	-	-.511	-	-	.687	-	-	*0.588*

Note. The lower diagonals contain posterior means, the upper diagonals posterior credibility intervals for the latent correlations. Latent factor variances are given in italics on the diagonals. CMS_{Hl}: SHS common method state residual variables for measurement occasion l; CMS_{Sl}: SWLS common method state residual variables for measurement occasion l; CMT_{Hi}: SHS common method trait variable for indicator i; CMT_{Si}: SWLS common method trait variable for indicator i; MS_{Hl}: SHS method state residual variables for measurement occasion l; MS_{Sl}: SWLS method state residual variables for measurement occasion l; MT_{Hi}: SHS method trait variable for indicator i; MT_{Si}: SWLS method trait variable for indicator i; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale.

Table 8.12: Latent variance coefficients for the SWLS and SHS items in the LST-Com GRM: Consistency and method specificity coefficients on the trait and occasion-specific levels

	SWLS									SHS								
	item 1			item 2			item 3			item 1			item 2			item 3		
Coeff.	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
Parents																		
TMS	.624	.624	.624	.702	.702	.702	.567	.567	.567	.671	.671	.671	.691	.691	.691	.665	.665	.665
	[.504, .739]	[.504, .739]	[.504, .739]	[.587, .808]	[.587, .808]	[.587, .808]	[.463, .671]	[.463, .671]	[.463, .671]	[.555, .781]	[.555, .781]	[.555, .781]	[.584, .793]	[.584, .793]	[.584, .793]	[.557, .770]	[.557, .770]	[.557, .770]
TCon	.376	.376	.376	.298	.298	.298	.433	.433	.433	.329	.329	.329	.309	.309	.309	.335	.335	.335
	[.261, .496]	[.261, .496]	[.261, .496]	[.192, .413]	[.192, .413]	[.192, .413]	[.329, .537]	[.329, .537]	[.329, .537]	[.219, .445]	[.219, .445]	[.219, .445]	[.207, .416]	[.207, .416]	[.207, .416]	[.230, .443]	[.230, .443]	[.230, .443]
Peers																		
TCMS	.175	.175	.175	.188	.188	.188	.202	.202	.202	.230	.230	.230	.227	.227	.227	.193	.193	.193
	[.047, .301]	[.047, .301]	[.047, .301]	[.086, .305]	[.086, .305]	[.086, .305]	[.089, .323]	[.089, .323]	[.089, .323]	[.110, .358]	[.110, .358]	[.110, .358]	[.086, .359]	[.086, .359]	[.086, .359]	[.086, .312]	[.086, .312]	[.086, .312]
TUMS	.614	.614	.614	.668	.668	.668	.561	.561	.561	.586	.586	.586	.602	.602	.602	.607	.607	.607
	[.490, .731]	[.490, .731]	[.490, .731]	[.550, .777]	[.550, .777]	[.550, .777]	[.441, .683]	[.441, .683]	[.441, .683]	[.461, .716]	[.461, .716]	[.461, .716]	[.473, .746]	[.473, .746]	[.473, .746]	[.489, .725]	[.489, .725]	[.489, .725]
TCon	.210	.210	.210	.144	.144	.144	.238	.238	.238	.184	.184	.184	.171	.171	.171	.200	.200	.200
	[.139, .289]	[.139, .289]	[.139, .289]	[.085, .214]	[.085, .214]	[.085, .214]	[.168, .314]	[.168, .314]	[.168, .314]	[.116, .259]	[.116, .259]	[.116, .259]	[.109, .238]	[.109, .238]	[.109, .238]	[.133, .272]	[.133, .272]	[.133, .272]
TCon	.554	.554	.554	.443	.443	.443	.549	.549	.549	.451	.451	.451	.440	.440	.440	.516	.516	.516
(level 2)	[.354, .759]	[.354, .759]	[.354, .759]	[.253, .664]	[.253, .664]	[.253, .664]	[.377, .743]	[.377, .743]	[.377, .743]	[.277, .656]	[.277, .656]	[.277, .656]	[.266, .685]	[.266, .685]	[.266, .685]	[.337, .720]	[.337, .720]	[.337, .720]
Parents																		
OMS	.968	.935	.938	.993	.986	.986	.949	.898	.901	.984	.987	.982	.939	.946	.930	.917	.930	.906
	[.897, 1.000]	[.790,.999]	[.806, .999]	[.965, 1.000]	[.925, 1.000]	[.930, 1.000]	[.879, .991]	[.759, .983]	[.767, .984]	[.921, 1.00]	[.930, 1.00]	[.909, 1.00]	[.817, .999]	[.808, .999]	[.791, .999]	[.715, 1.00]	[.755, 1.00]	[.690, .999]
OCon	.032	.065	.062	.007	.014	.014	.051	.102	.099	.016	.013	.018	.061	.054	.070	.083	.070	.094
	[.000, .103]	[.001, .210]	[.001, .194]	[.000, .035]	[.000, .075]	[.000, .070]	[.009, .121]	[.017, .241]	[.016, .233]	[.000, .079]	[.000, .007]	[.000, .091]	[.001, .183]	[.001, .182]	[.001, .209]	[.000, .285]	[.000, .245]	[.001, .031]
Peers																		
OCMS	.489	.222	.267	.229	.090	.109	.149	.051	.066	.360	.307	.405	.028	.022	.032	.076	.064	.089
	[.257, .699]	[.096, .409]	[.098, .504]	[.010, .517]	[.003, .293]	[.001, .031]	[.003, .345]	[.004, .168]	[.004, .217]	[.180, .575]	[.145, .513]	[.202, .622]	[.000, .144]	[.000, .108	[.000, .149]	[.000, .324]	[.000, .281]	[.000, .354]
OUMS	.485	.744	.695	.734	.873	.847	.826	.924	.905	.621	.679	.576	.966	.974	.961	.907	.924	.892
	[.280, .714]	[.551, .881]	[.458, .874]	[.450, .959]	[.660, .980]	[.604, .974]	[.593, .966]	[.798, .985]	[.747, .981]	[.408, .804]	[.472, .844]	[.359, .782]	[.847, .999]	[.887, .999]	[.841, .999]	[.657, .998]	[.704, .998]	[.622, .997]
OCon	.026	.034	.039	.037	.038	.044	.026	.025	.029	.019	.014	.020	.006	.004	.007	.017	.012	.019
	[.001, .075]	[.001, .097]	[.001, .109]	[.001, .113]	[.001, .115]	[.001, .132]	[.001, .078]	[.001, .076]	[.001, .084]	[.000, .080]	[.000, .061]	[.000, .081]	[.000, .030]	[.000, .022]	[.000, .035]	[.000, .080]	[.000, .059]	[.000, .091]

Note. Credibility intervals are given in parentheses below the posterior mean values. Coeff.: Coefficient; OCMS: Occasion-specific common method state specificity; OCon: Occasion-specific state consistency; OMS: Occasion-specific method state specificity; OUMS: Occasion-specific unique method state specificity; SHS: Subjective Happiness Scale; SWLS: Satisfaction with Life Scale; Tl: measurement occasion *l*; TCMS: Common method trait specificity; TCon: Trait consistency; TCon (level 2): Trait Consistency on the between-level for the interchangeable peer reports; TMS: Method trait specificity (structurally different parent reports); TUMS: Unique method trait specificity. Coefficients were calculated using the formulas given in Table 4.1.

CMT correlations for the SHS items having CIs that include zero). Occasion-specific, transient over- or underestimation of the target's happiness or life satisfaction is not shared by the peers (as a group) and the parents (see Table 8.11).
Estimates of trait- as well as occasion-specific consistency and method specificity coefficients along with their 95% CI for the items of the SWLS and SHS are presented in Table 8.12.
Consistency coefficients on the level of stable interindividual differences in the parent reports range from .298 to .433. That is, depending on the item, between 29.8% and 43.4% of the reliable and stable variance in the SWLS parent reports can be explained by stable differences in the SWLS self-reports. Trait consistency coefficients of the SHS parent report items are minimally smaller, where 30.9% to 33.5% of the stable variance of the latent response variable is shared with the stable self-report latent state variables. The largest amount of the stable variance in the peer reports is variance that goes back to the stable individual views of the peers, not shared with either the self-report or the other peers. This is the case for both the SWLS and the SHS, as indicated by unique method trait specificity coefficients of .561 - .668 (SWLS) and .586 - .607 (SHS). In contrast, a stable common view of the peers that is not shared with the stable self-report accounts for 17.5% - 20.2% of the stable interindividual differences in the SWLS peer reports and for 19.3% - 23.0% of the stable variance in the SHS peer reports. Trait consistency coefficients for the peer reports (.144 - .238) are smaller than those for the parent reports. However, if only the common view of the peers is considered, i.e., disregarding interindividual differences between the different peers and taking the expected value over interchangeable peer ratings per target, trait consistency rises to .440 - .554.
Regarding occasion-specific, transient rater effects, convergence between parent or peer ratings and self-ratings drop off to almost zero. That is, occasion-specific consistencies for SWLS and SHS parent as well as peer reports range between .004 and .102. Hence, occasion-specific views are highly rater-dependent. Also on the occasion-specific level, the largest amount of variance in the peer reports is variance explained by the transient individual views of the specific peer raters, that is not shared with the other peers.

8.4.2 Discussion

The good model fit of the LST-Com GRM shows that the observed indicator-specific stabilities in the SWLS and SHS measures are well recovered by the LST-Com GRM covariance structure.
Both SWLS and SHS self-report measures as well as informant report measures were observed to be highly stable, with stable indicator-specific effects, however still indicating some variability in both components over time.
The application of the LST-Com GRM to the SWB data allows to answer several research questions:

- Does global life satisfaction (as the cognitive component of SWB) or global subjective happiness (as the affective part of SWB) show a greater stability over time?
- How stable are rater-effects over time?
- Is the association between life satisfaction and subjective happiness larger on a stable (trait-) level or on a transient, time-specific level?
- Does the degree of rater-convergence differ for cognitive and affective parts of SWB?
- Can rater-effects be generalized over the two SWB components?
- Does the degree of rater-convergence differ between stable and transient components of SWB?
- Are there item-specific effects in the measurement, stability and rater-bias for the SWLS and SHS items?

The results for the LST-Com GRM presented above give some valuable insights into the stability of SWLS and SHS measures, different rater effects and the stability and generalizability of these rater effects.

Both life satisfaction and subjective happiness themselves were found to be quite stable, neither of the two components being clearly more or less stable than the other.

Associations between life satisfaction and subjective happiness are high on both the stable (trait-)level as well as on the momentary, occasion-specific level. Hence, targets that show a higher stable level of life satisfaction also tend to have a higher stable level of happiness and vice versa. Furthermore, targets that deviate from their habitual level of life satisfaction into a positive (negative) direction tend to also show higher (lower) levels of subjective happiness than habitually at a respective measurement time point (and vice versa).

Self-informant correlations were comparable for life satisfaction and subjective happiness as well as the different items of the scales, consistent with the results of the meta-analysis by L. Schneider and Schimmack (2009). Therefore, peer ratings or parent ratings do not seem to be less or more biased (with reference to the self-report) for one of the two well-being components.

Self-informant correlations on a time-stable level observed in the current analysis ranged between .546 and .658 for the parents and between .663 and .744 for the peers. That is, trait-level, measurement error-free self-informant correlations are larger than the typically found average self-informant correlation of .42 reported by L. Schneider and Schimmack (2009).

Observed rater effects seem to be partly generalizable over the two components on the trait-level. For the peers, this is not only the case for individual raters but also for the common view of the peers on the target. Rater-effects on the occasion-specific level, however, are only observed to be associated between the two well-being components for the parent ratings. In none of the cases, generalizability across the two well-being components was high enough to argue for a common method effect. That is, the need for construct-specific method effects is clearly given.

Peers and parents share a common view, not shared with the targets' self-reports, regarding the stable, enduring life-satisfaction, however not with regard to happiness or occasion-specific life-satisfaction or happiness. In general, occasion-specific views are highly rater-dependent, meaning that the targets' perceived transient fluctuations around their habitual trait-level for both life satisfaction and happiness are not "correctly" perceived by the peers or the parents.

On the stable trait-level, convergent validity is higher for the target and the parent ratings as compared to the convergent validity between the target- and an individual-peer rating. That is, if the interest lies in approximating the targets' self-perceived happiness or life satisfaction as close as possible by an informant rating and only one rater was available for an informant report rating, it would be preferable to choose a parent instead of a peer. However, if the ratings of several interchangeable peers are available for a target, the common view of the peers shows higher convergent validity with the self-report than the parent reports do for both life satisfaction and happiness.

With respect to the item-analyses, the better fit of the model with indicator-specific state variables (tested for the LS-Com GRM) as compared to common latent state variables is not surprising, as items (or even test-halfs) of a scale are rarely perfectly homogeneous. The analysis on the item-level revealed that Item 2 of the SWLS seems to reflect an aspect of life satisfaction that is in parts distinct to the aspects measured by Item 1 and Item 3 and also shows less associations with subjective happiness. This is consistent with theoretical considerations on item formulation, in that Item 2 covers the part of life satisfaction that refers to comparatively more objective criteria, explicitly referring to the conditions of one's life instead of, e.g., satisfaction with the former. Furthermore, the results indicate that in general the answer scale is not used in different ways by the different rater-groups, but that parents tend to answer slightly more positive to the SWLS items than targets and peers do.

8.5 LGC-Com GRM of Life Satisfaction and Subjective Happiness

In this section the LGC-Com GRM is applied to the SWLS and SHS self-, parent and peer report measures. The specific LGC-Com GRM used is the latent growth curve variant of the LST-Com GRM simulated in Section 7.7 and corresponds to the model depicted in Figure 7.7, with the difference that the model in this application includes three methods instead of two (as depicted in the Figure), that is, an additional structurally different method (the parent reports). That is, the LGC model applied to the data assumed trait change only in the reference-method indicators. Due to the results of the application of the LS-Com GRM and LST-Com GRM to these data (i.e., indicator-specific effects and stabilities), the LGC-Com GRM was specified with indicator-specific latent intercept and slope variables. As MI was tested for the LS-Com GRM and LST-Com GRM in Sections 8.3 and 8.4 and the LGC-Com GRM represents a restrictive variant of the LST-Com change GRM, MI testing was not repeated here. As reported in aforementioned sections, the requirement of strong MI for the latent growth curve model is fulfilled. The LGC-Com GRM was estimated assuming a linear growth trajectory, with freely estimated latent slope means.

8.5.1 Results

Model fit results for the LGC-Com GRM are given in Table 8.13. The model fits the data reasonably well, with a PPP-value of .225 and a 95% confidence interval for the difference between the observed and replicated Chi-Square values including zero, [-102.20, 233.73]. However, traceplots indicated non-convergence of the loadings of the slope factors on the non-reference method (parent and peer report) items. Furthermore, also PSR values indicated suboptimal convergence despite the large number of iterations. Estimated latent means of the SWLS and SHS slope factors range from 0.115 to 0.233 and from 0.022 to 0.087, respectively. That is, a small uniform trait change for all targets / raters, i.e., a change in the latent trait means, can be observed for the SWLS items but is close to zero for the SHS items. The variance estimates of the slope factors are small, ranging from 0.066 to 0.112 for the SWLS slope factors and from 0.084 to 0.098 for the SHS slope factors. These small growth factor variances indicate that there are little to no inter-individual differences in intra-individual growth over time.

Due to the lack of convergence and of substantial variance in intra-individual growth over time, no variance coefficients will be reported for the LGC-Com GRM. All other parameter estimates are similar to those obtained in the LST-Com GRM application. Consequently, the LST-Com GRM seems to be more appropriate for the current data. For results and interpretations refer to Section 8.4.

Considering the results of the simulation study on the LGC-Com GRM (see Section 7.7), it is not surprising to encounter convergence problems for the slope factor loadings in the current application. That is, empirical identification of the slope factors can become difficult and estimation can run into problems if growth factor variances are small. This effect can be observed especially in cases where the number of measurement occasions or sample size are suboptimal as well (Bishop et al., 2015). As the present study includes only three measurement occasions and less observations than recommended for the application of the LGC-Com GRM by results of the simulation study (i.e., at least 600 between-level and 10 within-level observations), conditions for estimating the LGC-Com GRM are suboptimal.

Table 8.13: Model fit results for the LGC-Com GRM of life satisfaction and subjective happiness

MLR estimator (indicators treated as continuous)									
Remarks	χ^2	df	RMSEA	CFI	SRMR W	SRMR B	AIC	BIC	BIC adj.
Ψ [a]	1799.74	1380	0.015	0.973	0.040	0.063	51551.48	53279.00	52230.72
Bayesian estimation (indicators treated as categorical)									
			χ^2-diff. limit						
Traceplots	PSR	PPP	lower	upper	Iterations	Thinning	Chains		
problematic[b]	1.053	.225	-102.20	233.73	200,000	10	3		

Note. Growth is restricted to be linear, reference-method indicator slope loadings are set to 1 for T2 and 2 for T3; AIC: Akaike information criterion; BIC: Bayesian information criterion; BIC adj.: sample size adjusted BIC; CFI: Comparative fit index; MLR: Maximum likelihood robust estimator (Mplus); PPP: Posterior predictive p-value; Ψ: The latent variance-covariance matrix Ψ is not positive definite due to negative variance estimates; PSR: Potential scale reduction factor; RMSEA: Root mean square error of approximation; SRMR B: standardized root mean square residual on the between-level; SRMR W: standardized root mean square residual on the within-level.

[a] Negative variance estimates for one SWLS and two SHS Slope factors.

[b] Traceplots look good for all parameters except for the loadings of the slope factors on the non-reference method items, which indicate a lack of convergence.

8.6 Discussion of the model applications

This chapter presented an application of the LS-Com GRM, LST-Com GRM, and LGC-Com GRM to self-report, parent report and friend-report data of SWB for recent high-school graduates in Germany. On the basis of these life satisfaction and subjective happiness ratings, it was illustrated how the models can be used to analyze convergent and discriminant validity over time, analyze change and stability of construct and method effects over time, investigate the generalizability of method effects across methods or time, and use item-specific effects (e.g., correlations or method-specific difficulties) for item-selection or rater-selection.

Furthermore, it was illustrated that Bayesian methods can provide credibility intervals for key quantities of the models such as method specificities or construct and method stabilities. Classical "frequentist" confidence intervals wich are based on normal theory may not be trustworthy for these types of parameters that are likely to have skewed distributions.

The vast majority of previous studies on SWB based stability and rater-consistency estimates on correlations of observed variables. This approach is problematic as measurement error, on the one hand, and instability and rater-inconsistencies, on the other hand, are confounded and might thus produce biased estimates of stabilities and consistencies (Eid & Diener, 2004). The present application investigated inter-rater consistencies and stability of rater-effects for life satisfaction or subjective happiness while explicitly taking measurement error into account.

Time-stable self-informant correlations observed in the LST-Com GRM were larger than the typically found average self-informant correlation of .42 reported by L. Schneider and Schimmack (2009). That is, convergent validity was observed to be higher than in previous studies. This might not only be due to the fact that correlations were based on measurement-error free latent variables, but primarily to the observation that self-informant correlation on the occasion-specific level are close to zero while being large on the stable trait-level. Models that do not separate time-stable and occasion-specific, momentary components of SWB measures might find reduced correlations due to the negligibly small rater-convergences on the occasion-specific level. In general, occasion-specific views were observed to be highly rater-dependent, meaning that the targets' perceived transient fluctuations around their habitual trait-level for both life satisfaction and happiness are not shared with the peers' or the parents' perception. These results stress the importance of separating long-term differences in SWB from occasion-specific effects. Furthermore, as noted by Eid and Diener (2004), not short-term fluctuations of mood and emotions but rather more stable aspects of life are of primary interest in quality of life research.

Observed rater effects seem to be partly generalizable over the two components on the trait-level. However, in none of the cases generalizability across the two well-being components was high enough to argue for a common method factor. That is, the need for construct-specific method effects is clearly given.

Additionally, the results underline that it is important to consider multiple methods when assessing subjective well-being, as the findings suggest that each rater group had a specific perspective on the targets' life satisfaction and happiness. Nevertheless, the model results can give hints with regard to rater selection for informant ratings on SWB measures. On the stable trait-level, convergent validity was higher for the target- and the parent ratings as compared to the convergent validity between the target and an individual peer rating. That is, if only one rater was available for an informant report rating, it seems advisable to choose a parent instead of a peer in order to yield a rating closer to the target's own perception. If the ratings of several peers per target are merged, these compound measures show higher convergent validity with the self-reports than the parent reports do. However, this conclusion is only justified if the targets' perspective is of primary interest. As ratings are always rater-specific there is not one "correct" rating, but only a perspective that might be considered most relevant or most appropriate for the (research) question at hand.

Life satisfaction and subjective happiness were both found to be highly stable. The results of applying the LGC-Com GRM to the present SWB data suggest that inter-individual differences in intra-individual change are, if present, small. As the estimation of the model exhibited convergence problems, the possibility that such differences exist cannot be precluded. However, the good fit of the LST-Com GRM with strong MI is an additional indication that growth variances might be negligible. Note that by establishing strong MI, the specified LST-Com GRM is a pure state-variability model. That is, this model does not include any form of trait change, but assumes that state scores fluctuate around an invariant, stable set-point.

Chapter 9

Final Discussion

9.1 Summary and Conclusions

In the present work, different longitudinal MTMM graded response model for the combination of structurally different and interchangeable methods were introduced. The models combine the advantages of multilevel MTMM measurement designs and longitudinal CFA models for categorical indicators. Thus far, only few models have been presented allowing researchers to analyze MTMM data with ordered response variables (Crayen et al., 2011; Eid, 1996; Jeon & Rijmen, 2014; Nussbeck et al., 2006). However, none of these models can be used for longitudinal MTMM measurement designs combining structurally different and interchangeable methods. The presented models fill this gap in the current literature on longitudinal MTMM modeling.

While previous studies had to focus on models including only structurally different methods or use aggregated scores for the interchangeable methods in order to analyze different types of methods in one model, the presented GRMs overcome these limitations. Furthermore, the models allow to disentangle different sources of variance and thereby separate trait and method components. Additionally, the impact of method effects can be analyzed on both levels of measurement. Stability and change of inter-individual differences in an attribute as well as of method factors can be computed and the generalizability of method effects can be analyzed. Furthermore, the models avoid an undesirable loss of information at the item-level caused by the aggregation of items in order to obtain continuous outcomes. In the models presented here, convergent and discriminant validity can be computed on the item-level. This information might be useful, for instance, for test construction. Furthermore, it allows to estimate CFA-MTMM models even if only relatively few items per construct are available. Until now, only Bayesian data analysis renders the estimation of these complex models possible. Additionally, Bayesian sampling offers a range of further possibilities and advantages. First, it allows researchers to estimate credibility intervals for parameters with unknown and potentially skewed distributions. This is especially relevant for CFA-MTMM analyses where variance components and correlations are used as indicators for convergent and discriminant validity, method specificities or stabilities. Second, Bayesian estimation methods have been found to exhibit better small-sample performances for factor analyses (Depaoli & Clifton, 2015; Lee & Song, 2004) or when there are only few clusters in multilevel models (Asparouhov & Muthén, 2010b; Hox et al., 2012).

As the results of our simulation study show, the LS-Com GRM and LST-Com GRM can be accurately estimated under realistic sample sizes if low degrees of convergent validity are present. These results are encouraging and suggest that even complex multilevel longitudinal CFA-MTMM models can be applied in a wide range of situations using Bayesian methods.

However, with only few level-1 units (e.g., 2 observations per cluster) convergence problems become more likely and bias levels increase. The same holds for parameters associated with factors with low

variances, e.g., the method factors under high consistency or the slope factors in the LGC-Com GRM. The possibility to incorporate informative prior information in the estimation process increases the applicability of the LS-Com GRM in small samples, does, however, not significantly benefit estimation accuracy in the LST-Com and LGC-Com GRMs.
An application of the models to multi-rater data on life satisfaction and subjective happiness illustrated the applicability and advantages of the models in applied research. The results underline the importance of considering several different methods in research on subjective well-being. Furthermore, the results stress the importance of separating long-term differences in SWB from occasion-specific effects and investigate rater-effects and rater-convergence on both the stable and the occasion-specific level.

9.2 Recommendations for Applied researchers

Model choice. The longitudinal models presented in this work differ in their assumptions made about the underlying change process and thereby also in the research questions they are suitable to answer. While in the LS-Com GRM, the degree of stability in constructs and method effects can be investigated by means of latent correlations, intraindividual change is not directly modeled or represented by a latent variable. If the interest lies in investigating change more explicitly, the LC-Com or LGC-Com GRM would be a more appropriate choice. The LS-Com and LC-Com GRMs, on the other hand, do not differentiate between state variability and trait change. If a researcher is interested in separating stable inter-individual differences from occasion-specific variability, the LST-Com GRM or the LGC-Com GRM as defined in the previous chapters are the models of choice. These models allow to analyze true discriminant and convergent validity on the level of occasion-specific as well as on the level of stable variables.
Another decision that depends on the research question at hand is the choice of a reference method. The reference method should be selected based on theoretical considerations and the contrast between methods that is most meaningful with respect the research question. The choice of a reference method is not restricted to either structurally different or the set of interchangeable methods. For guidelines as to the choice of the reference method see Geiser et al. (2008), for an example on how to use the set of interchangeable methods as reference method see Pham et al. (2012).
Applied researchers should bear in mind that including additional constructs may increase the complexity of the model and thereby also its estimation. For researchers wishing to include a larger number of constructs in their analysis (e.g., more than two), it might be an option to split the model into several submodels and analyze all possible combinations of two constructs per model separately (Koch, 2013). This approach is a valid option in the application of either of the models, as it does not change the meaning of the latent variables or of the coefficients of convergent and discriminant validity, stabilities or method generalizabilities.
Furthermore, model complexity should be reduced where possible and indicated. For instance, many of the permissible covariances between latent factors in the presented models can be set to zero for parsimony reasons, as they can be expected to be non-significant in many empirical applications.

Sample size. Based on the results of the simulation study, researchers are recommended to sample a minimum number of 250 targets and 5 raters per target in order to obtain reliable parameter estimates in the LS-Com GRM. If informative prior information is incorporated in the analysis, fewer numbers of level-1 units may be sufficient (> 2). In contrast, it is recommended to sample at least 500 targets with a minimum of 5 raters per target when applying the LST-Com GRM. Additionally, it is recommendable to include at least three measurement occasions (and more than one construct) in applications of the LST-Com GRM.

For the estimation of the LGC-Com GRM sample sizes of at least 600 between-level and 10 within-level observations should be included in the analyses. However, whenever slope variances are estimated to be very small and the estimation shows convergence problems for the respective slope parameters in applications of the LGC-Com GRM, it is advisable to resort to the comparably less complex LST-Com GRM, as the underlying process might be one of state variability around stable inter-individual differences.

Use of prior information. The use of informative priors did not reduce the number of required observations on the between- or the within-level in the LST-Com GRM and LGC-Com GRM. Increasing the degree of prior informativeness could potentially reduce the sample sizes needed for accurate parameter estimation. This approach does, however, involve the risk of detrimental effects caused by setting incorrect prior locations (see, e.g., Holtmann et al., 2016). Consequently, highly informative priors should be employed with caution. Whenever researchers wish to use informative priors, it is recommend to conduct sensitivity analyses to scrutinize the impact of prior assumptions on parameter estimates. In addition, in applied Bayesian analyses, the inspection of convergence by use of visual diagnostic tools such as trace plots (Gelman et al., 2014; Lynch, 2007) is indispensable.

Measurement invariance. Researchers interested in mean change of a construct over time in the LS-Com or LC-Com GRM should ensure that strong measurement invariance holds (Meredith & Teresi, 2006). Also when researchers wish to model autoregressive effects on the state residual variables in the LST-Com or LGC-Com GRM it is recommended to establish MI of the state residual factor loading parameters to ensure that the autoregressive effects can be meaningfully interpreted.
However, partial measurement invariance (Byrne, Shavelson, & Muthén, 1989) might be sufficient under certain circumstances. Steenkamp and Baumgartner (1998) suggested that equality of at least two factor loadings and intercepts (in the case of continuous indicators) is sufficient to interpret latent mean differences. However, as discussed in the respective sections on measurement invariance, researchers should keep in mind their measurement invariance settings when interpreting latent mean differences.
Measurement invariance could also be tested across methods (Geiser, Burns, & Servera, 2014), in order to investigate latent mean differences in ratings across different rater groups. For instance, the question whether the answer scale of the SWLS and SHS items is used in different ways by the targets, the parents, and the peers and whether there are mean differences in their ratings, could have been investigated by testing MI of the threshold parameters across the rater groups. However, testing nested models against each other using Bayesian data analysis is still challenging due to the lack of an adequate difference test and the limitation of the PPP that less restrictive models will always show a better PPP value. This issue might be resolved by using Bayesian estimation programs that provide the deviance information criterion (DIC) in their output or by comparing parameters via credibility intervals. The latter approach gets cumbersome, however, if many items with a large number of categories are to be estimated. As in the case of a large number of constructs it might be an option to split the model into sub-models in order to test for MI using WLSMV estimation where possible, or to use classical model fit results of MLR estimation (e.g., RMSEA, CFI) as rough indicators of model fit. Furthermore, it has been argued that, for instance, the magnitude of intercept differences is of greater importance than the statistical significance of a difference (B. Muthén & Asparouhov, 2012; Steinmetz, 2013). An approach explicitly tackling this criticism is the concept of approximate MI (B. Muthén & Asparouhov, 2013; van de Schoot et al., 2013), which builds on Bayesian estimation and replaces exact zero by approximate zero constraints by setting informative, small-variance priors on parameter differences. The option to set priors on parameter differences in Mplus is currently allowed for intercept, slope and loading parameters, however, has not yet been implemented for thresholds of polytomous items in Mplus 7.3 (L. K. Muthén & Muthén, 1998-2012).

Explaining method effects. Often researchers are interested in identifying predictors or consequences of trait and method effects. When relating external variables to these factors in the presented CFA-MTMM models several caveats have to be considered. In a recent article, Koch, Holtmann, Bohn, and Eid (in press) have shown that external variables cannot be directly related to the latent factors in g-factor type of models using a classical multiple-indicator multiple-cause (MIMIC) approach, as this would violate some of the basic psychometric properties of the latent variables and thereby lead to model misspecification and parameter bias. More specifically, using one or more external variables as regression predictors for the general (e.g., state) or specific (e.g., method) factors leads to a violation of the basic property that general and specific factors are uncorrelated by definition if the external variables are also correlated with the respective other factor. Additionally, when explaining method factors, the psychometric property that residual factors always have a mean of zero has to be considered. To overcome these methodological problems, researchers need to transform the explanatory variables before using them as predictors for the latent factors in g-factor models. Koch, Holtmann, et al. (in press) proposed two possible modeling strategies to circumvent these problems and give detailed guidelines with respect to their application.
In contrast, using external variables as outcome variables or merely using external variables as correlates (no directional regression approach) is unproblematic in terms of a violation of the psychometric properties. However, researchers have to keep in mind that when correlating external variables with factors in CTC(M-1) type of models the associations will depend on the choice of the reference method. That is, correlations of external variables with the common method variables (on the between-level) represent semi-partial correlations, in which the non-reference method is corrected for influences of the reference method (Geiser et al., 2008, 2012).

9.3 Limitations and Future Directions

As discussed previously, it would have been interesting to investigate the impact of incorrectly specified informative priors on estimation accuracy. The impact of priors is especially important as the choice of an appropriate prior mean can be challenging in practice and setting inaccurate priors might have detrimental effects on parameter estimates especially in categorical indicator models, due to a phenomenon termed prior assumption dependence (Asparouhov & Muthén, 2010b; Depaoli & Clifton, 2015). So far, only little is known about the effect of setting priors with inaccurate locations and further research is needed. The inclusion of several different correctly and incorrectly specified informative prior conditions in the simulation design had to be discarded in the present work given the complexity of the models and the resulting simulation time. The precise investigation of the influence of inaccurate priors is a task that might be better investigated using a simpler model and that has for instance been addressed for a cross-sectional multilevel MTMM model including interchangeable raters only (Holtmann et al., 2016). In general, due to the growing use of measurement designs requiring longitudinal multilevel MTMM GRMs and their estimation by Bayesian methods, additional research is needed on the effect informative priors have on estimation accuracy in multilevel GRMs.
Similarly, the effect of the number of response categories on estimation accuracy and its potential interactions with sample size or cell frequencies is an interesting question that should be pursued in future research. As the effects of increasing the number of response categories on parameter estimation has been found to be negligible to positive (e.g., Flora & Curran, 2004; Forero & Maydeu-Olivares, 2009; Liang & Yang, 2014; Li, 2016; Moshagen & Musch, 2014), the recommended sample sizes should be sufficient to estimate the corresponding model for observed variables with more than three response categories, too.
Furthermore, the hypothesis that estimation problems of the LGC-Com GRM slope parameters might

vanish in models with more substantial slope variances should be investigated in future studies within a smaller design.

In the context of MTMM analysis, the possibility to include prior information in the estimation process seems especially interesting with respect to the parameters of convergent and discriminant validity. This is not only the case as they belong to the parameters of primary interest in MTMM studies, but also as they might be parameters that are most widely reported in the previous literature. However, variance components and correlation coefficients are defined as functions of primary model parameters estimated in the presented CFA-MTMM models. To the best of my knowledge, Mplus 7.3 does not allow researchers to set priors on secondary, newly defined model parameters, such as parameter ratios. Instead, priors have to be set on the respective variance and covariance parameters in order to indirectly influence estimates of variance components. This approach is not only less than ideal, but defining priors for variance and covariance parameters is also challenging, especially if variance covariance matrices grow large and Inverse Wishart priors get difficult to handle. This is due to the property of the Inverse Wishart distribution that the informativeness of one parameter in the matrix determines the informativeness of other parameters (Asparouhov & Muthén, 2010b; Tokuda et al., 2011). Inverse Wishart priors may therefore not be sufficiently flexible for some applications. Alternative prior settings for variance and covariance parameters, such as, e.g., decomposing covariance matrices into a scale and a correlation matrix (Lewandowski et al., 2009; Stan Development Team, 2014b) and their applicability to the CFA-MTMM coefficients should be explored in future studies. The possibility of this decomposition and the resulting option to set priors on correlation matrices might render this approach especially interesting in the context of the presented models.

In this context, the use of the software Mplus might be mentioned as another limitation. In comparison to other software programs for MCMC estimation, such as Jags (Plummer, 2003) or Stan (Stan Development Team, 2014a), Mplus offers less flexibility in prior settings and parameterizations in the model specification. However, Mplus is one of the most widely applied programs for structural equation modeling and has the advantage that it is comparably fast in its MCMC estimation and specifically designed to estimate SEMs. In contrast, Stan, for instance, is an open-source software coded in C++, offering the possibilities of more flexible modeling. Additionally Stan is not restricted to the use of conjugate priors (Stan Development Team, 2014b) and has a wider variety of possible prior distributions that can be assigned to the parameters. Hence, it is worth investigating the possibilities other software programs offer for the estimation of longitudinal CFA-MTMM models. On the other hand, it has to be kept in mind that the complexity of the presented models renders the specification of these models in programs such as Jags or Stan rather complicated and error-prone, such that Mplus remains the recommended option for applied researchers.

Regarding the simulation study, Mplus was chosen as the preferred program as the long simulation times excluded the use of programs with potentially even longer estimation times. As implementations of MCMC methods and algorithms vary widely across software packages, estimation results might differ between software packages. It can therefore not be precluded that estimation results would look slightly different if the stimulation studies were conducted with a different software program. However, for the estimation of a cross-sectional multilevel CFA-MTMM model for interchangeable raters the same number of observations were needed to obtain accurate parameter estimates in both Mplus and Stan (Holtmann et al., 2016). Potential differences in estimation results are therefore expected to be minimal.

In general, further extensions of the models presented in this work are not directly considered for future research, as estimation as well as applicability already reach their limits due to model complexity and large required sample sizes. However, one interesting extension would be to incorporate the possibility of modeling individually varying times of observations, allowing for variable lags between observations over time and across individuals. A clear advantage of this extension is that it would reduce the importance of choosing an appropriate lag for investigating stability and thereby ease the comparison of effects across studies. Individually varying time lags could be incorporated into the models by specifying autoregressive or growth curve loading parameters as a function of the lag length (Eid et al., 2012) or by combining the presented modeling approaches with recently developed continuous time models (Oud, 2002; Oud & Delsing, 2010; Voelkle, Oud, Davidov, & Schmidt, 2012).

References

Adler, M. G., & Fagley, N. S. (2005). Appreciation: Individual differences in finding value and meaning as a unique predictor of subjective well-being. *Journal of Personality*, *73*, 79–114. doi: 10.1111/j.1467-6494.2004.00305.x

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Asparouhov, T., & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *Proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology* (pp. 2531–2535).

Asparouhov, T., & Muthén, B. (2010a). *Bayesian analysis in Mplus: Technical implementation* (Mplus technical report, Version 3). Retrieved from `http://www.statmodel.com/download/Bayes3.pdf`

Asparouhov, T., & Muthén, B. (2010b). *Bayesian analysis of latent variable models using Mplus* (Technical report, Version 4). Retrieved from `www.statmodel.com/download/BayesAdvantages18.pdf`

Asparouhov, T., & Muthén, B. (2012). *General random effect latent variable modeling: Random subjects, items, contexts, and parameters.* Retrieved from `https://www.statmodel.com/download/NCME12.pdf`

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151–164. doi: 10.1037/a0030642

Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385–426). Greenwich, CT: Information Age.

Bauer, H. (2002). *Wahrscheinlichkeitstheorie* [Probability Theory] (5th ed.). New York, NY: de Gruyter.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*, 186–203. doi: 10.1207/s15328007sem1302_2

Béguin, A. A., & Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, *66*, 541–561. doi: 10.1007/BF02296195

Bishop, J., Geiser, C., & Cole, D. A. (2015). Modeling latent growth with multiple indicators: A comparison of three approaches. *Psychological Methods*, *20*, 43–62. doi: 10.1037/met0000018

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research*, *32*, 336–383. doi: 10.1177/0049124103260222

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.

Boomsma, A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 518–540. doi: 10.1080/10705511.2013.797839

Boyce, C. J., Wood, A. M., & Brown, G. D. (2010). The dark side of conscientiousness: Conscientious people experience greater drops in life satisfaction following unemployment. *Journal of Research in Personality*, *44*, 535–539. doi: 10.1016/j.jrp.2010.05.001

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466. doi: 10.1037/0033-2909.105.3.456

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. doi: 10.1037/h0046016

Carretero-Dios, H., Eid, M., & Ruch, W. (2011). Analyzing multitrait-mulitmethod data with multilevel confirmatory factor analysis: An application to the validation of the state-trait cheerfulness inventory. *Journal of Research in Personality*, *45*, 153–164. doi: 10.1016/j.jrp.2010.12.007

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research*, *29*, 468–508. doi: 10.1177/0049124101029004003

Cole, D. A. (2006). Coping with longitudinal data in research on developmental psychopathology. *International Journal of Behavioral Development*, *30*, 20-25. doi: 10.1177/0165025406059969

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods*, *10*, 3–20. doi: 10.1037/1082-989X.10.1.3

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design and statistical model. *Annual Review of Psychology*, *57*, 505–528. doi: 10.1146/annurev.psych.57.102904.190146

Collins, L. M., & Flaherty, B. P. (2002). Latent class models for longitudinal data. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 287–303). Cambridge, England: Cambridge University Press.

Courvoisier, D. S. (2006). *Unfolding the constituents of psychological scores: Development and application of mixture and multitrait-multimethod LST models*. Unpublished doctoral dissertation, Université de Genève, Geneva, Switzerland. Retrieved from `https://doc.rero.ch/record/9221/files/these.pdf`

Courvoisier, D. S., Eid, M., Lischetzke, T., & Schreiber, W. H. (2010). Psychometric properties of a computerized mobile phone method for assessing mood in daily life. *Emotion*, *10*, 115–124. doi: 10.1037/a0017813

Courvoisier, D. S., Nussbeck, F. W., Eid, M., Geiser, C., & Cole, D. A. (2008). Analyzing the convergent and discriminant validity of states and traits: Development and applications of multimethod latent state-trait models. *Psychological Assessment*, *20*, 270–280. doi: 10.1037/a0012812

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. London, UK: Chapman & Hall.

Crayen, C., Geiser, C., Scheithauer, H., & Eid, M. (2011). Evaluating interventions with multimethod data: A structural equation modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 497–524. doi: 10.1080/10705511.2011.607068

de Finetti, B. (1974). *Theory of probability*. Chichester: Wiley.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*, 197–229.

Depaoli, S. (2014). The impact of inaccurate ”informative” priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 239–252. doi: 10.1080/10705511.2014.882686

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 327–351. doi: 10.1080/10705511.2014.937849

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*, 542–575. doi: 10.1037/0033-2909.95.3.542

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, *3*, 1–43. doi: 10.1111/j.1758-0854.2010.01045.x

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71–75. doi: 10.1207/s15327752jpa4901_13

Diener, E., Ng, W., Harter, J., & Arora, R. (2010). Wealth and happiness across the world: Material prosperity predicts life evaluation whereas psychosocial prosperity predicts positive feeling. *Journal of Personality and Social Psychology*, *99*, 52–61. doi: 10.1037/a0018066

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309–326. doi: 10.1111/j.2044-8317.1994.tb01039.x

Dunson, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, *98*, 555–563. doi: 10.1198/016214503000000387

Edwards, P., Roberts, I., Sandercock, P., & Frost, C. (2004). Follow-up by mail in clinical trials: Does questionnaire length matter? *Controlled Clinical Trials*, *25*, 31–52. doi: 10.1016/j.cct.2003.08.013

Eid, M. (1995). *Modelle der Messung von Personen in Situationen* [Models for measuring persons in situations]. Weinheim, Germany: Psychologie Verlags Union.

Eid, M. (1996). Longitudinal confirmatory factor analysis for polytomous item responses: Model definition and model selection on the basis of stochastic measurement theory. *Methods of Psychological Research Online*, *1*(4), 65–85. Retrieved from `http://www.dgps.de/fachgruppen/methoden/mpr-online/issue1/art4/eid.pdf`

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261. doi: 10.1007/BF02294377

Eid, M. (2006). Latent class models for analyzing variability and change. In A. Ong & M. van Dulmen (Eds.), *Oxford handbook of methods in positive psychology* (pp. 591–608). Oxford, England: Oxford University Press.

Eid, M., Courvoisier, D. S., & Lischetzke, T. (2012). Structural equation modeling of ambulatory assessment data. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 384–406). New York, NY: Guilford Press.

Eid, M., & Diener, E. (2004). Global judgments of subjective well-being: Situational variability and long-term stability. *Social Indicators Research*, *65*, 245–277. doi: 10.1023/B:SOCI.0000003801.89195.bc

Eid, M., & Diener, E. (2006). *Handbook of multimethod measurement in psychology.* Washington, DC: American Psychological Association.

Eid, M., Geiser, C., & Koch, T. (2016). Measuring method effects from traditional to design-oriented approaches. *Current Directions in Psychological Science*, *25*, 275–280. doi: 10.1177/0963721416649624

Eid, M., & Hoffmann, L. (1998). Measuring variability and change with an item response model for polytomous variables. *Journal of Educational and Behavioral Statistics*, *23*, 193–215. doi: 10.3102/10769986023003193

Eid, M., Holtmann, J., Santangelo, P., & Ebner-Priemer, U. (in press). On the definition of latent-state-trait models with autoregressive effects: Insights from LST-R theory. *European Journal of Psychological Assessment*.

Eid, M., & Kutscher, T. (2014). Statistical models for analyzing stability and change in happiness. In K. M. Sheldon & R. E. Lucas (Eds.), *Stability of happiness: Theories and evidence on whether happiness can change* (pp. 261–297). London, England:

Elsevier.

Eid, M., & Langeheine, R. (1999). The measurement of consistency and occasion specificity with latent class models: A new model and its application to the measurement of affect. *Psychological Methods*, *4*, 100–116. doi: 10.1037/1082-989X.4.1.100

Eid, M., & Langeheine, R. (2003). Separating stable from variable individuals in longitudinal studies by mixture distribution models. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 179–206. doi: 10.1207/S15366359MEA0103_01

Eid, M., Lischetzke, T., Nussbeck, F., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, *8*, 38–60. doi: 10.1037/1082-989X.8.1.38

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, *13*, 230–253. doi: 10.1037/a0013219

Eid, M., Schneider, C., & Schwenkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality*, *13*, 283–306.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology*, *4*, 22–36. doi: 10.1027/1614-2241.4.1.22

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research*, *37*, 379–403. doi: 10.1207/S15327906MBR3703_4

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, *112*, 393–395. doi: 10.1037/0033-2909.112.3.393

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491. doi: 10.1037/1082-989X.9.4.466

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: Limited versus full information methods. *Psychological Methods*, *14*, 275–299. doi: 10.1037/a0015825

Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous items. *British Journal of Mathematical and Statistical Psychology*, *58*, 145–172. doi: 10.1348/000711005X38951

Fox, J.-P., & Glas, C. A. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288. doi: 10.1007/BF02294839

Geiser, C. (2008). *Structural equation modeling of multitrait-multimethod-*

multioccasion data. Unpublished doctoral dissertation, Freie Universität Berlin, Germany. Retrieved from `http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000004268/Dissertation_Christian_Geiser.pdf?hosts=`

Geiser, C., Bishop, J., & Lockhart, G. (2015). Collapsing factors in multitrait-multimethod models: examining consequences of a mismatch between measurement design and model. *Frontiers in Psychology*, *6*, 946. doi: 10.3389/fpsyg.2015.00946

Geiser, C., Burns, G. L., & Servera, M. (2014). Testing for measurement invariance and latent mean differences across methods: Interesting incremental information from multitrait-multimethod studies. *Frontiers in Psychology*, *5*, 1216. doi: 10.3389/fpsyg.2014.01216

Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, *13*, 49–57. doi: 10.1037/1082-989X.13.1.49

Geiser, C., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Multitrait-multimethod change modelling. *Advances in Statistical Analysis*, *94*, 185–201. doi: 10.1007/s10182-010-0127-0

Geiser, C., Eid, M., West, S. G., Lischetzke, T., & Nussbeck, F. W. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 409–436. doi: 10.1080/10705511.2012.687658

Geiser, C., Keller, B. T., & Lockhart, G. (2013). First- versus second-order latent growth curve models: Some insights from latent state-trait theory. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*, 479–503. doi: 10.1080/10705511.2013.797832

Geiser, C., Keller, B. T., Lockhart, G., Eid, M., Cole, D. A., & Koch, T. (2015). Distinguishing state variability from trait change in longitudinal data: The role of measurement (non) invariance in latent state-trait analyses. *Behavior Research Methods*, *47*, 172–203. doi: 10.3758/s13428-014-0457-z

Geiser, C., Koch, T., & Eid, M. (2014). Data-generating mechanisms versus constructively defined latent variables in multitrait–multimethod analysis: A comment on castro-schilo, widaman, and grimm (2013). *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 509–523. doi: 10.1080/10705511.2014.919816

Geiser, C., & Lockhart, G. (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods*, *17*, 255–283. doi: 10.1037/a0026977

Gelman, A. (2013). Two simple examples for understanding posterior p-values whose distributions are far from uniform. *Electronic Journal of Statistics*, *7*, 2595–2602. doi: 10.1214/13-EJS854

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall / CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–472. doi: 10.1214/ss/1177011136

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 721–741. doi: 10.1109/TPAMI.1984.4767596

Glaesmer, H., Grande, G., Braehler, E., & Roth, M. (2011). The German version of the Satisfaction With Life Scale (SWLS). *European Journal of Psychological Assessment*, *27*, 127–132. doi: 10.1027/1015-5759/a000058

Grimm, K. J., Pianta, R. C., & Konold, T. (2009). Longitudinal multitrait-multimethod models for developmental research. *Multivariate Behavioral Research*, *44*, 233–258. doi: 10.1080/00273170902794230

Hallquist, M., & Wiley, J. (2014). MplusAutomation: Automating Mplus model estimation and interpretation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=MplusAutomation` (R package version 0.6-3)

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*, 404–416. doi: 10.1177/0049124104270220

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. doi: 10.1037/a0038889

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. (2007). The integrated trait–state model. *Journal of Research in Personality*, *41*, 295–315. doi: 10.1016/j.jrp.2006.04.003

Headey, B., Muffels, R., & Wagner, G. G. (2013). Choices which change life satisfaction: Similar results for Australia, Britain and Germany. *Social Indicators Research*, *112*, 725–748. doi: 10.1007/s11205-012-0079-8

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS*. New York, NY: Routledge.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait-state distinction for the structural model of developmental change. *Child Development*, *58*, 93–109.

Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 541–563. doi: 10.1080/10705510802338983

Hjort, N. L., Dahl, F. A., & Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *Journal of the American Statistical Association*, *101*, 1157–1174. doi: 10

.1198/016214505000001393

Hojtink, H., & Molenaar, I. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika*, *62*, 171–189. doi: 10.1007/BF02295273

Holtmann, J., Koch, T., Bohn, J., & Eid, M. (2017). Bayesian analysis of longitudinal mulitrait-multimethod data with ordinal response variables. *British Journal of Mathematical and Statistical Psychology*, *70*, 42-80. doi: 10.1111/bmsp.12081

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study. *Multivariate Behavioral Research*, *51*, 661–680. doi: 10.1080/00273171.2016.1208074

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 157–174. doi: 10.1207/S15328007SEM0802_1

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, *6*, 87–93. doi: 10.18148/srm/2012.v6i2.5033

Jackman, S. (2009). *Bayesian analysis for the social sciences*. Chichester: John Wiley & Sons.

Jeon, M., & Rijmen, F. (2014). Recent developments in maximum likelihood estimation of MTMM models for categorical data. *Frontiers in Psychology*, *5*, 269. doi: 10.3389/fpsyg.2014.00269

Jovanovic, V. (2011). Personality and subjective well-being: One neglected model of personality and two forgotten aspects of subjective well-being. *Personality and Individual Differences*, *50*, 631–635. doi: 10.1016/j.paid.2010.12.008

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 325–352. doi: 10.1207/S15328007SEM0803_1

Kenny, D. A. (1995). The multitrait-multimethod matrix: Design, analysis, and conceptual issues. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 111–124). Hillsdale, NJ: Erlbaum.

Kenny, D. A., & Zautra, A. (2001). Trait–state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243–263). Washington, DC: American Psychological Association. doi: 10.1037/10409-008

Khoo, S. T., West, S. G., Wu, W., & Kwok, O. M. (2006). Longitudinal methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 301–317). Washington, DC: American Psychological Association.

Koch, T. (2013). *Multilevel structural equation modeling of multitrait-multimethod-*

multioccasion data. Unpublished doctoral dissertation, Freie Universität Berlin, Germany. Retrieved from http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000013668/Thesis_Koch_gedreht.pdf?hosts=

Koch, T., Eid, M., & Lochner, K. (in press). Multitrait-multimethod-analysis: The psychometric foundation of CFA-MTMM models. In P. Irwing, T. Booth, & D. Hughes (Eds.), *The Wiley handbook of psychometric testing.* London, England: Wiley.

Koch, T., Holtmann, J., Bohn, J., & Eid, M. (in press). Explaining general and specific factors in longitudinal, multimethod, and bi-factor models: Some caveats and recommendations. *Psychological Methods*.

Koch, T., Schultze, M., Eid, M., & Geiser, C. (2014). A longitudinal multilevel CFA-MTMM model for interchangeable and structurally different methods. *Frontiers in Psychology*, *5*, 311. doi: 10.3389/fpsyg.2014.00311

Koch, T., Schultze, M., Holtmann, J., Geiser, C., & Eid, M. (2017). A multimethod latent state-trait model for structurally different and interchangeable methods. *Psychometrika*, *82*, 17–47. doi: 10.1007/s11336-016-9541-x

Koch, T., Schultze, M., Jeon, M., Nussbeck, F., Praetorius, A.-K., & Eid, M. (2016). A cross-classified CFA-MTMM model for structurally different and nonindependent interchangeable methods. *Multivariate Behavioral Research*, *51*, 67–85. doi: 10.1080/00273171.2015.1101367

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London, England: Elsevier.

Kwok, O., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research*, *42*, 557–592. doi: 10.1080/00273170701540537

Lang, F., Weiss, D., Stocker, A., & von Rosenbladt, B. (2007). Assessing cognitive capacities in computer-assisted survey research: Two ultra-short tests of intellectual ability in the German Socio-Economic Panel (SOEP). *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, *127*, 183–192.

Langeheine, R. (1994). Latent variable markov models. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 373–395). Thousand Oaks, CA: Sage.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*, 653–686. doi: 10.1207/s15327906mbr3904_4

Lee, S.-Y., Song, X.-Y., & Cai, J.-H. (2010). A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 280–302. doi: 10.1080/10705511003659425

Leite, W. L. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 581–610. doi: 10.1080/10705510701575438

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*, 1989–2001. doi: 10.1016/j.jmva.2009.04.008

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihodd and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936–949. doi: 10.3758/s13428-015-0619-7

Liang, X., & Yang, Y. (2014). An evaluation of WLSMV and Bayesian methods for confirmatory factor analysis with categorical indicators. *International Journal of Quantitative Research in Education*, *2*, 17–38. doi: 10.1504/IJQRE.2014.060972

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*, 112–115. doi: 10.1111/j.2041-210X.2011.00131.x

Little, T. D., Schnabel, K. U., & Baumert, J. (2000). *Modeling longitudinal and multilevel data.* Mahwah, NJ: Erlbaum.

Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology*, *71*, 616–628. doi: 10.1037/0022-3514.71.3.616

Lucas, R. E., & Donnellan, M. B. (2007). How stable is happiness? Using the STARTS model to estimate the stability of life satisfaction. *Journal of Research in Personality*, *41*, 1091–1098. doi: 10.1016/j.jrp.2006.11.005

Luhmann, M., Bohn, J., Holtmann, J., Koch, T., & Eid, M. (2016). I'm lonely, can't you tell? Convergent validity of self- and informant ratings of loneliness. *Journal of Research in Personality*, *61*, 50–60. doi: 10.1016/j.jrp.2016.02.002

Luhmann, M., Hawkley, L. C., Eid, M., & Cacioppo, J. T. (2012). Time frames and the distinction between affective and cognitive well-being. *Journal of Research in Personality*, *46*, 431–441. doi: 10.1016/j.jrp.2012.04.004

Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: A meta-analysis on differences between cognitive and affective well-being. *Journal of Personality and Social Psychology*, *102*, 592–615. doi: 10.1037/a0025948

Luhmann, M., Schimmack, U., & Eid, M. (2011). Stability and variability in the relationship between subjective well-being and income. *Journal of Research in Personality*, *45*, 186–197. doi: 10.1016/j.jrp.2011.01.004

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. doi: 10.1007/978-0-387-71265-9.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137–155. doi: 10.1023/A:1006824100041

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86–92. doi: 10.1027/1614-1881.1.3.86

Mahlke, J., Schultze, M., Koch, T., Eid, M., Eckert, R., & Brodbeck, F. (2016). A multilevel CFA-MTMM approach for multisource feedback instruments: Presentation and application of a new statistical model. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 91–110. doi: 10.1080/10705511.2014.990153

Marsh, H. W. (1993). Multitrait-multimethod analyses: Inferring each trait/method combination with multiple indicators. *Applied Measurement in Education*, *6*, 49-81. doi: 10.1207/s15324818ame0601_4

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmators factors analysis. *Journal of Applied Psychology*, *73*, 107–117. doi: 10.1037/0021-9010.73.1.107

Mayer, A., Geiser, C., Infurna, F. J., & Fiege, C. (2013). Modelling and predicting complex patterns of change using growth component models: An application to depression trajectories in cancer patients. *European Journal of Developmental Psychology*, *10*, 40–59. doi: 10.1080/17405629.2012.732721

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). doi: 10.1007/978-1-4613-0893-5_17

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133. doi: 10.2307/1130295

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). doi: 10.1037/10409-005

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In W. F. Velicer & J. A. Schinka (Eds.), *Handbook of psychology: Research methods in psychology* (pp. 447—480). New York, NY: Wiley.

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics*, *22*, 1142–1160. doi: 10.1214/aos/1176325622

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. doi: 10.1007/BF02294825

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11), 69–77. doi: 10.1097/01.mlr.0000245438.73837.89

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, *3*, 45–58. doi: 10.18148/srm/2009.v3i1.666

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479–515. doi: 10.1207/S15327906MBR3903_4

Miyazaki, K., & Hoshino, T. (2009). A Bayesian semiparametric item response model with Dirichlet process priors. *Psychometrika*, *74*, 375–393. doi: 10.1007/s11336-008-9108-6

Moshagen, M., & Musch, J. (2014). Sample size requirements of the robust weighted least squares estimator. *Methodology*, *10*, 60–70. doi: 10.1027/1614-2241/a000068

Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Incomplete draft, Version 3). Retrieved from `http://www.statmodel.com/download/IntroBayesVersion%203.pdf`

Muthén, B., & Asparouhov, T. (2002). *Latent variable analysis with categorical outcomes: Multiple-Group and growth modeling in Mplus* (Mplus Web Notes Nos. 4, Version 5). Retrieved from `https://www.statmodel.com/download/webnotes/CatMGLong.pdf`

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. doi: 10.1037/a0026802

Muthén, B., & Asparouhov, T. (2013). *BSEM measurement invariance analysis* (Mplus Web Notes No. 17). Retrieved from `https://www.statmodel.com/examples/webnotes/webnote17.pdf`

Muthén, L. K., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 599–620. doi: 10.1207/S15328007SEM0904_8

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide.* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual changes. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (p. 92-105). Washington, DC: American Psychological Association.

Nesselroade, J. R., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*, 9–29. doi: 10.1080/15427609.2004.9683328

Newsom, J. T. (2015). *Longitudinal structural equation modeling. A comprehensive introduction*. New York: Routledge.

Nussbeck, F. W., Eid, M., & Lischetzke, T. (2006). Analysing multitrait–multimethod data with structural equation models for ordinal variables applying the WLSMV estimator: What sample size is needed for valid results? *British Journal of Mathematical and Statistical Psychology*, *59*, 195–213. doi: 10.1348/000711005X67490

Oud, J. H. L. (2002). Continuous time modeling of the cross-lagges panel design. *Kwantitatieve Methoden*, *69*, 1–26.

Oud, J. H. L., & Delsing, M. J. M. H. (2010). Continuous time modeling of panel data by means of SEM. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 201–244). Heidelberg, Germany: Springer.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov Chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*, 146–178. doi: 10.3102/10769986024002146

Pavot, W., & Diener, E. (1993a). The affective and cognitive context of self-reported measures of subjetive well-being. *Social Indicators Research*, *28*, 1–20. doi: 10.1007/BF01086714

Pavot, W., & Diener, E. (1993b). Review of the Satisfaction With Life Scale. *Psychological Assessment*, *5*, 164–172. doi: 10.1037/1040-3590.5.2.164

Pavot, W., & Diener, E. (2008). The Satisfaction With Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, *3*, 137–152. doi: 10.1080/17439760701756946

Pham, G., Koch, T., Helmke, A., Schrader, F.-W., Helmke, T., & Eid, M. (2012). Do teachers know how their teaching is perceived by their pupils? *Procedia - Social and Behavioral Sciences*, *46*, 3368–3374. doi: 10.1016/j.sbspro.2012.06.068

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*. Retrieved from `https://www.r-project.org/conferences/DSC-2003/Drafts/Plummer.pdf`.

Pohl, S., & Steyer, R. (2010). Modeling common traits and method effects in multitrait-multimethod analysis. *Multivariate Behavioral Research*, *45*, 45–72. doi: 10.1080/00273170903504729

Pohl, S., Steyer, R., & Kraus, K. (2008). Modelling method effects as individual causal effects. *Journal of the Royal Statistical Society: Series A*, *171*, 41–63. doi: 10.1111/j.1467-985X.2007.00517.x

Prenoveau, J. M. (2016). Specifying and interpreting latent state-trait models with autoregression: An illustration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 731–749. doi: 10.1080/10705511.2016.1186550

R Development Core Team. (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from `http://www.r-project.org`

Rabe-Hesketh, S., & Skrondal, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190. doi: 10.1007/BF02295939

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete variable models with nested random effects. *Journal of Econometrics*, *128*, 301-323. doi: 10.1016/j.jeconom.2004.08.017

Rammstedt, B., & Beierlein, C. (2015). Can't we make it any shorter? *Journal of Individual Differences*, *35*, 212–220. doi: 10.1027/1614-0001/a000141

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*, 354–373. doi: 10.1037/a0029315

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25. doi: 10.1037/0033-2909.126.1.3

Rovine, M. J., & Walls, T. A. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 124–147). New York, NY: Oxford University Press.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*, 1–97. doi: 10.1007/BF03372160

Samejima, F. (2010). The general graded response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 77–108). New York, NY: Taylor & Francis.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179–200). doi: 10.1037/10409-006

Schermelleh-Engel, K., Keith, N., Moosbrugger, H., & Hodapp, V. (2004). Decomposing person and occasion–specific effects: An extension of latent state-trait (LST) theory to hierarchical LST models. *Psychological Methods*, *9*, 198–219. doi: 10.1037/1082-989X.9.2.198

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74. Retrieved from https://www.dgps.de/fachgruppen/methoden/mpr-online/issue20/art2/mpr130_13.pdf

Scherpenzeel, A., & Saris, W. (2007). Multitrait-multimethod models for longitudinal research. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 381–401). Mahwah, NJ: Lawrence Erlbaum.

Schiffrin, H. H., & Nelson, K. S. (2010). Stressed and happy? Investigating the relationship between happiness and perceived stress. *Journal of Happiness Studies*, *11*, 33–39. doi: 10.1007/s10902-008-9104-7

Schimmack, U., & Oishi, S. (2005). The influence of chronically and temporarily accessible information on life satisfaction judgments. *Journal of Personality and Social Psychology*, *89*, 395–406. doi: 10.1037/0022-3514.89.3.395

Schimmack, U., Schupp, J., & Wagner, G. G. (2008). The influence of environment and personality on the affective and cognitive component of subjective well-being. *Social Indicators Research*, *89*, 41–60. doi: 10.1007/s11205-007-9230-3

Schneider, L., & Schimmack, U. (2009). Self-informant agreement in well-being ratings: A meta-analysis. *Social Indicators Research*, *94*, 363–376. doi: 10.1007/s11205-009-9440-y

Schneider, S. M., & Schupp, J. (2014). Individual differences in social comparison and its consequences for life satisfaction: Introducing a short scale of the Iowa–Netherlands comparison orientation measure. *Social Indicators Research*, *115*, 767–789. doi: 10.1007/s11205-012-0227-1

Schumacher, J. (2003). SWLS - Satisfaction With Life Scale. In J. Schumacher, A. Klaiber, & E. Braehler (Eds.), *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden* (pp. 305–309). Göttingen, Germany: Hogrefe.

Seligman, M., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*, 5–14. doi: 10.1037/0003-066X.55.1.5

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Sivo, S., Fan, X., & Witta, L. (2005). The biasing effects of unmodeled ARMA time series processes on latent growth curve model estimates. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*, 215–231. doi: 10.1207/s15328007sem1202_2

Song, X.-Y., & Lee, S.-Y. (2012). A tutorial on the Bayesian approach for analyzing structural equation models. *Journal of Mathematical Psychology*, *56*, 135–148. doi: 10.1016/j.jmp.2012.02.001

Song, X.-Y., Lu, Z.-H., Hser, Y.-I., & Lee, S.-Y. (2011). A Bayesian approach for analyzing longitudinal structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 183–194. doi: 10.1080/10705511.2011.557331

Stan Development Team. (2014a). *Stan: A C++ library for probability and sampling*.

Version 2.2. Retrieved from `http://mc-stan.org/`

Stan Development Team. (2014b). Stan modeling language users guide and reference manual, version 2.5.0 [Computer software manual]. Retrieved from `http://mc-stan.org/documentation/`

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A*, *171*, 5–19. doi: 10.1111/j.1467-985X.2007.00509.x

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–107. doi: 10.1086/209528

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The Meaning in Life Questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, *53*, 80–93. doi: 10.1037/0022-0167.53.1.80

Steinmetz, H. (2013). Analyzing observed composite differences across groups. Is partial measurement invariance enough? *Methodology*, *9*, 1–12. doi: 10.1027/1614-2241/a000049

Steyer, R. (1988). Conditional expectations: An introduction to the concept and its applications in empirical sciences. *Methodika*, *2*, 53–78.

Steyer, R. (1989). Models of classical psychometric test theory as stochastic measurement models: Representation, uniqueness, meaningfulness, identifiability, and testability. *Methodika*, *3*, 25-60.

Steyer, R., & Eid, M. (2001). *Messen und Testen* [Measurement and testing.] (2nd ed.). Heidelberg, Germany: Springer.

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research Online*, *2*(1), 21–33.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*, 79-98.

Steyer, R., Majcen, A.-M., Schwenkmezger, P., & Buchner, A. (1989). A latent state-trait anxiety model and its application to determine consistency and specificity coefficients. *Anxiety Research*, *1*, 281–299. doi: 10.1080/08917778908248726

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits - revised. *Annual Review of Clinical Psychology*, *11*, 71–98. doi: 10.1146/annurev-clinpsy-032813-153719

Steyer, R., & Nagel, W. (2017). *Probability and conditional expectation: Fundamentals for the empirical sciences*. London, England: Wiley.

Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little & K. U. Schnabel (Eds.), *Modeling longitudinal and multiple-*

group data: Practical issues, applied approaches, and specific examples (pp. 109–126). Mahwah, NJ: Erlbaum.

Steyer, R., & Schmitt, M. (1990). The effects of aggregation across and within occasions on consistency, specificity and reliability. *Methodika*, *4*, 58–94. Retrieved from https://www.metheval.uni-jena.de/materialien/publikationen/Steyer,_R._and_Schmitt,_M.J.1990.pdf

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state-trait theory and research in personality and individual differences. *European Journal of Personality*, *13*, 389–408. doi: 10.1002/(SICI)1099-0984(199909/10)13:5\⟨389::AID-PER361\⟩3.0.CO;2-A

Swami, V., Stieger, S., Voracek, M., Dressler, S. G., Eisma, L., & Furnham, A. (2009). Psychometric evaluation of the Tagalog and German Subjective Happiness Scales and a cross-cultural comparison. *Social Indicators Research*, *93*, 393–406. doi: 10.1007/s11205-008-9331-7

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408. doi: 10.1007/BF02294363

Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods*, *5*, 175–198. doi: 10.1037/1082-989X.5.2.175

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., & Tuerlinckx, F. (2011). *Visualizing distributions of covariance matrices.* Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/Visualization.pdf

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770. doi: 10.3389/fpsyg.2013.00770

Vassar, M. (2008). A note on the score reliability for the Satisfaction With Life Scale: An RG study. *Social Indicators Research*, *86*, 47–57. doi: 10.1007/s11205-007-9113-7

Vautier, S. (2004). A longitudinal sem approach to STAI data: Two comprehensive multitrait-multistate models. *Journal of Personality Assessment*, *83*, 167–179. doi: 10.1207/s15327752jpa8302_11

Veenhoven, R. (1994). Is happiness a trait? Tests of the theory that a better society does not make people any happier. *Social Indicators Research*, *32*, 101–160. doi: 10.1007/1-4020-3742-2_17

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An sem approach to contiunous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*, 176–192. doi: 10.1037/a0027543

von Oertzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology*, *63*, 627–646. doi: 10.1348/000711010X486633

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel study (SOEP): Evolution, scope and enhancements. *Schmollers Jahrbuch: Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, *127*, 139-169. doi: 10.2139/ssrn.1028709

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 392–423. doi: 10.1080/10705511.2010.489003

Yuan, K.-H., & Hayashi, K. (2005). On Muthén's maximum likelihood for two-level covariance structure models. *Psychometrika*, *70*, 147–167. doi: 10.1007/s11336-003-1070-8

Zhang, J. W., Howell, R. T., & Stolarski, M. (2013). Comparing three methods to measure a balanced time perspective: The relationship between a balanced time perspective and subjective well-being. *Journal of Happiness Studies*, *14*, 169–184. doi: 10.1007/s10902-012-9322-x

Appendices

A Monte Carlo simulation study LS-Com GRM

A.1 Population parameters in the LS-Com GRM simulation

Table A 1: Population values in the LS-Com GRM Monte Carlo simulation study

Population parameters in the LS-Com GRM simulation						
	Low Con			High Con		
Parameter	$l = 2$	$l = 3$	$l = 4$	$l = 2$	$l = 3$	$l = 4$
Within-level						
UM						
Loadings	0.929 (± 0.026)	0.929 (±0.026)	0.929 (±0.025)	0.835 (±0.064)	0.835 (±0.063)	0.835 (±0.062)
Variances	0.275 (±0)	0.275 (±0)	0.275 (±0)	0.125 (±0)	0.125 (±0)	0.125 (±0)
Covariances	0.092 (±0.062)	0.090 (±0.060)	0.086 (±0.058)	0.042 (±0.028)	0.0408 (±0.027)	0.039 (±0.026)
Between-level						
State						
Loadings	0.752 (±0.189)	0.752 (±0.187)	0.752 (±0.186)	0.902 (±0.064)	0.902 (±0.063)	0.902 (±0.063)
Variances	0.825 (±0)	0.825 (±0)	0.825 (±0)	0.825 (±0)	0.825 (±0)	0.825 (±0)
Covariances	0.385 (±0.113)	0.358 (±0.123)	0.330 (±0.135)	0.385 (±0.113)	0.358 (±0.123)	0.330 (±0.135)
CM						
Loadings	1.080 (±0.028)	1.080 (±0.0269)	1.080 (±0.027)	1.223 (±0.073)	1.223 (±0.071)	1.223 (±0.070)
Variances	0.225 (±0)	0.225 (±0)	0.225 (±0)	0.075 (±0)	0.075 (±0)	0.075 (±0)
Covariances	0.075 (±0.051)	0.074 (±0.049)	0.071 (±0.047)	0.025 (±0.017)	0.025 (±0.016)	0.033 (±0.013)

Note. Mean values of the population parameters over all parameters of the same parameter type in the respective condition. Standard deviations of the different population values of the parameters of the respective parameter type are given in parentheses. A zero in the column *SD* indicates that the population value did not vary over the parameters of the respective parameter type. *CM*: common method factors; High Con; high consistency condition; l: number of measurement occasions; Low Con: Low consistency condition; *UM*: unique method factors.

A.2 Simulation results LS-Com GRM. Case of one construct ($j = 1$)

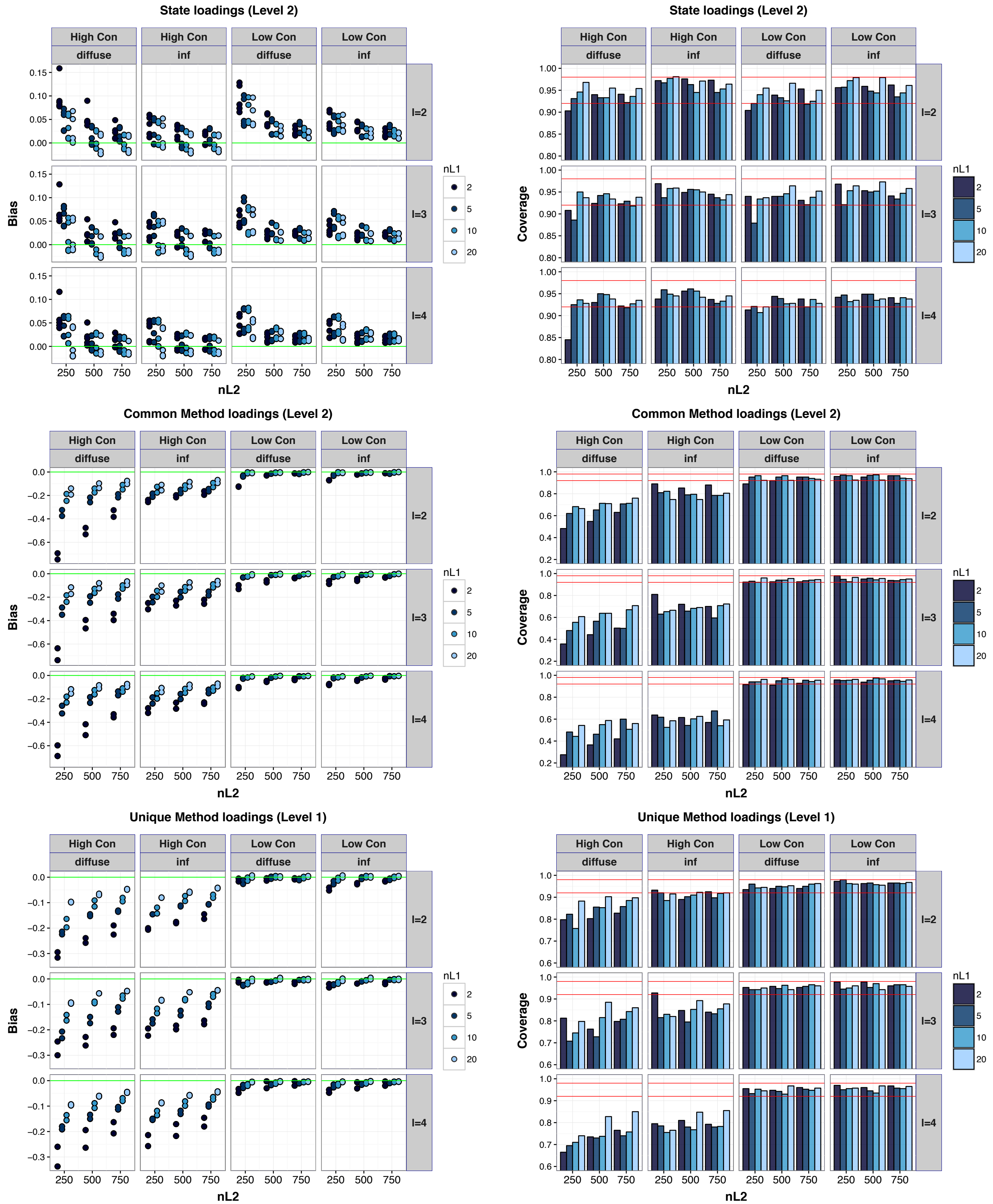

Figure A 1: Bias and 95% coverage for loading parameters in the LS-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

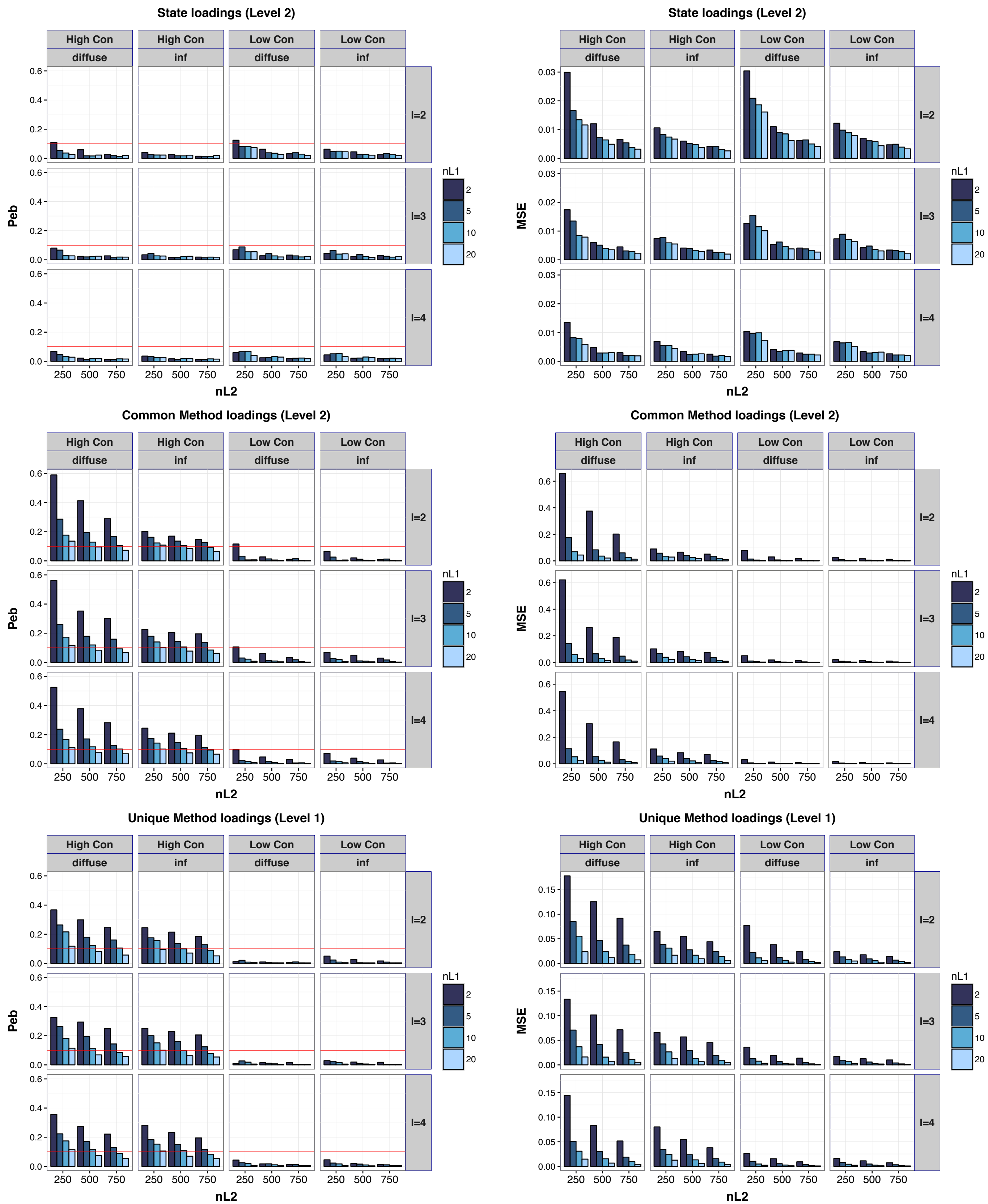

Figure A 2: Parameter estimation bias (peb) and mean squared error (MSE) for loading parameters in the LS-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

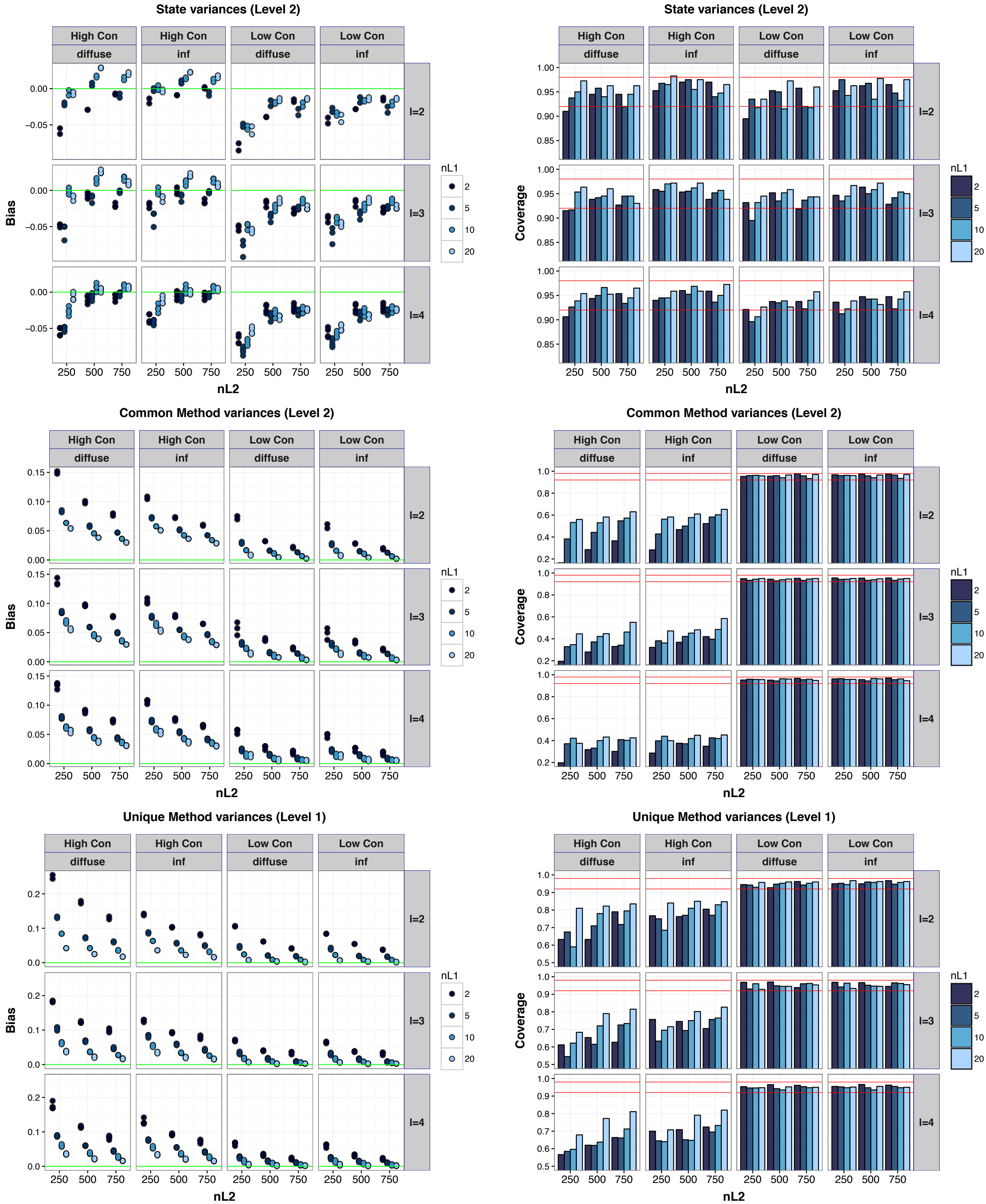

Figure A 3: Bias and 95% coverage for variance parameters in the LS-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

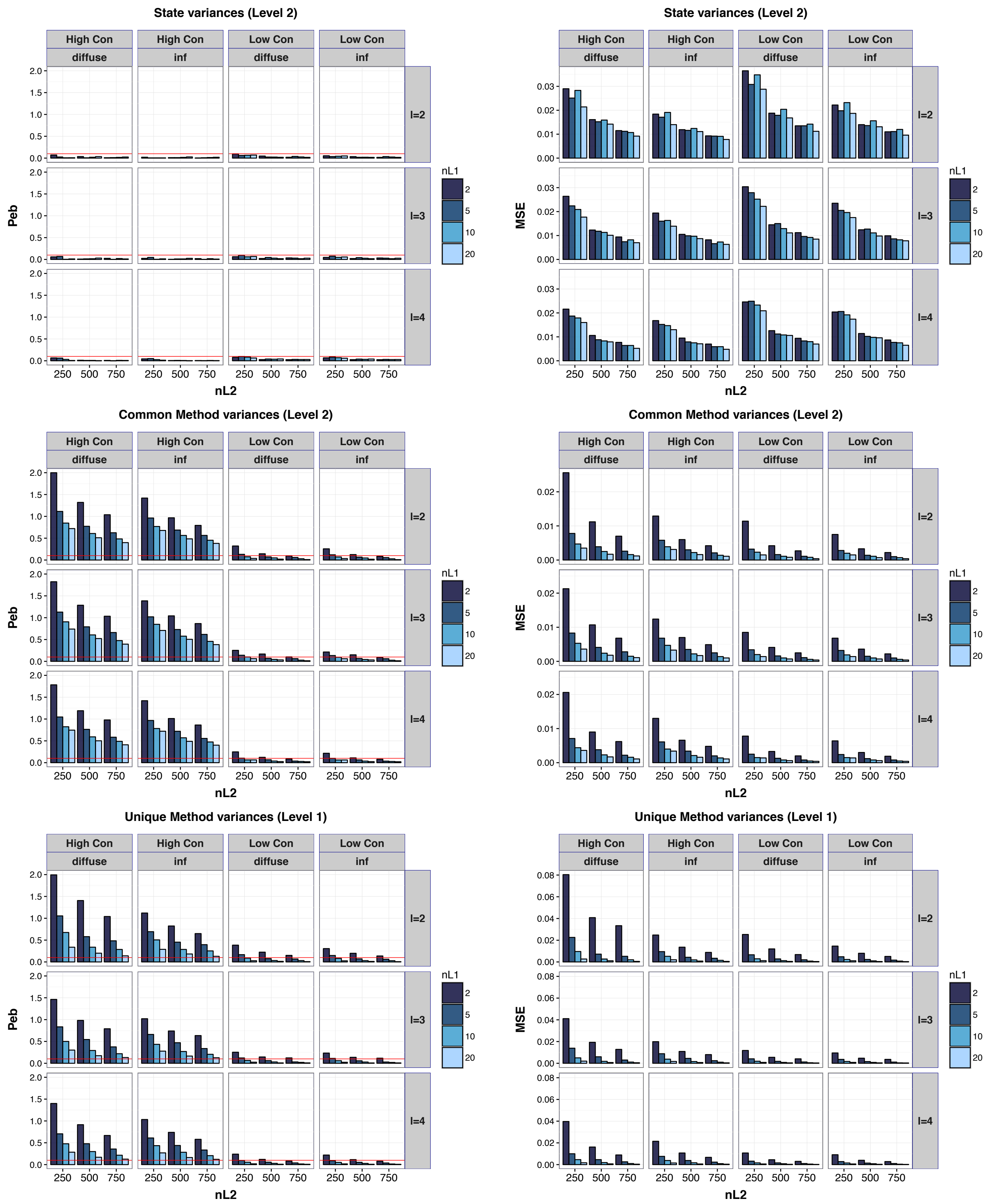

Figure A 4: Parameter estimation bias (peb) and mean squared error (MSE) for variance parameters in the LS-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations..

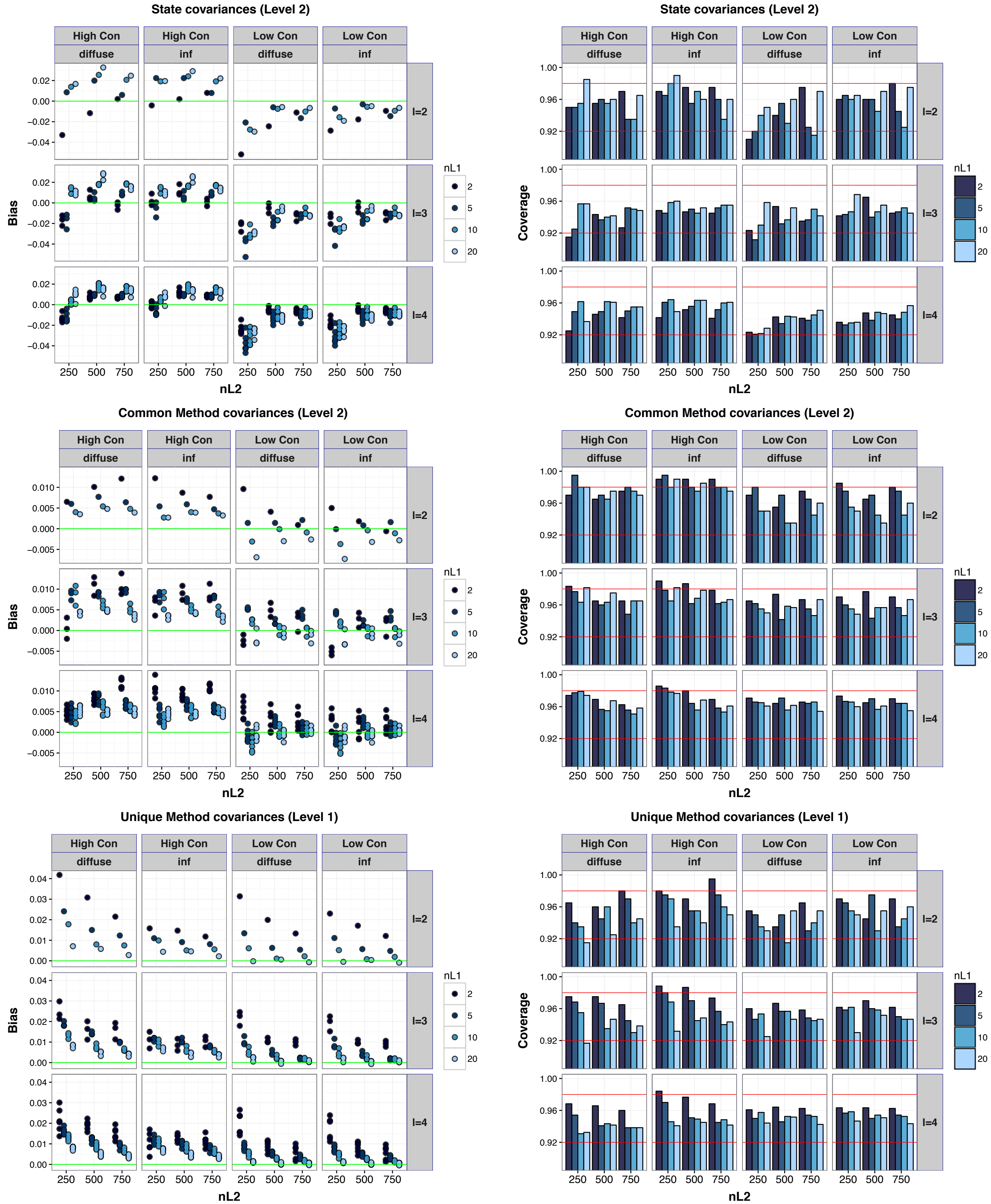

Figure A 5: Bias and 95% coverage for covariance parameters in the LS-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

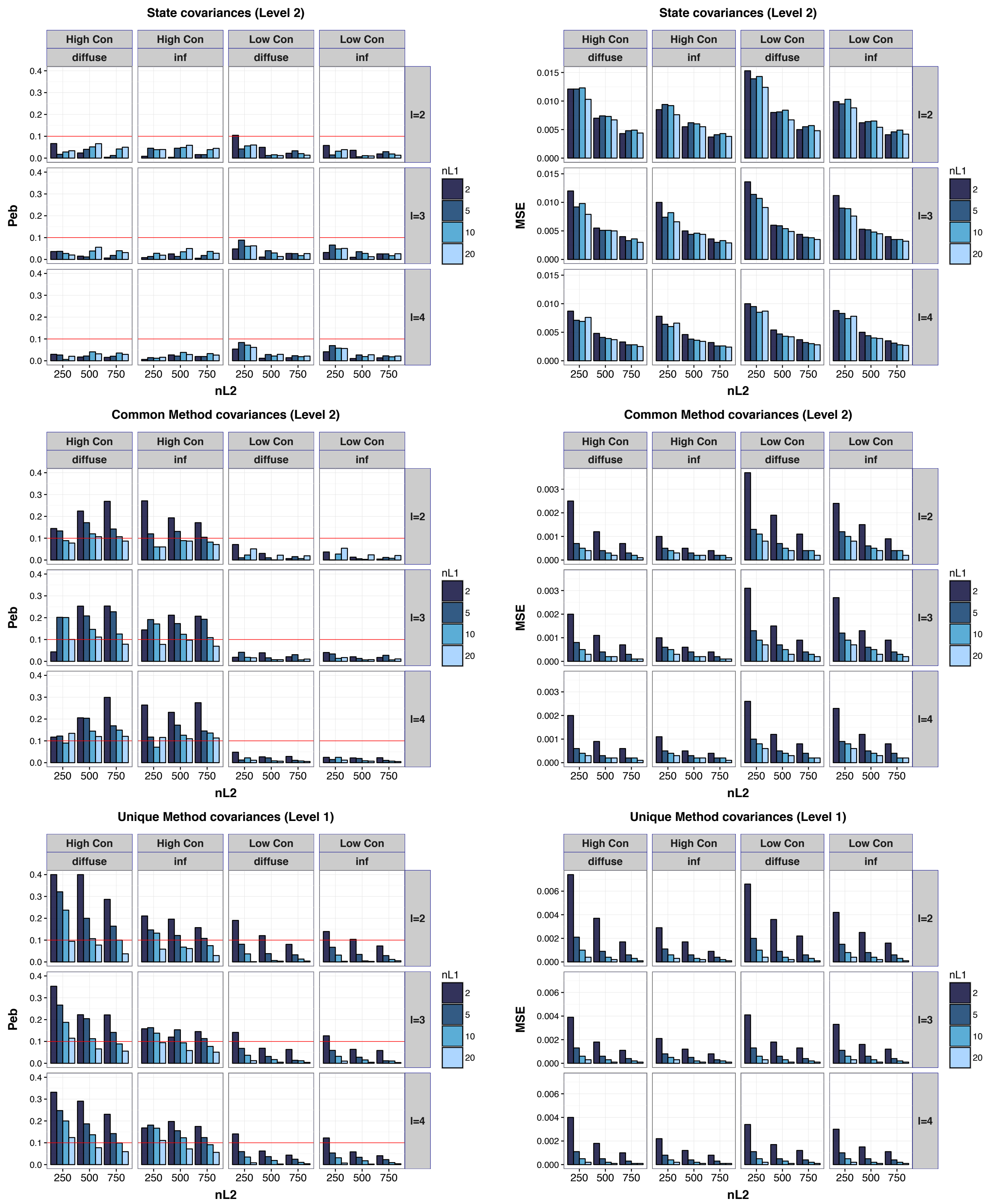

Figure A 6: Parameter estimation bias (peb) and mean squared error (MSE) for covariance parameters in the LS-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

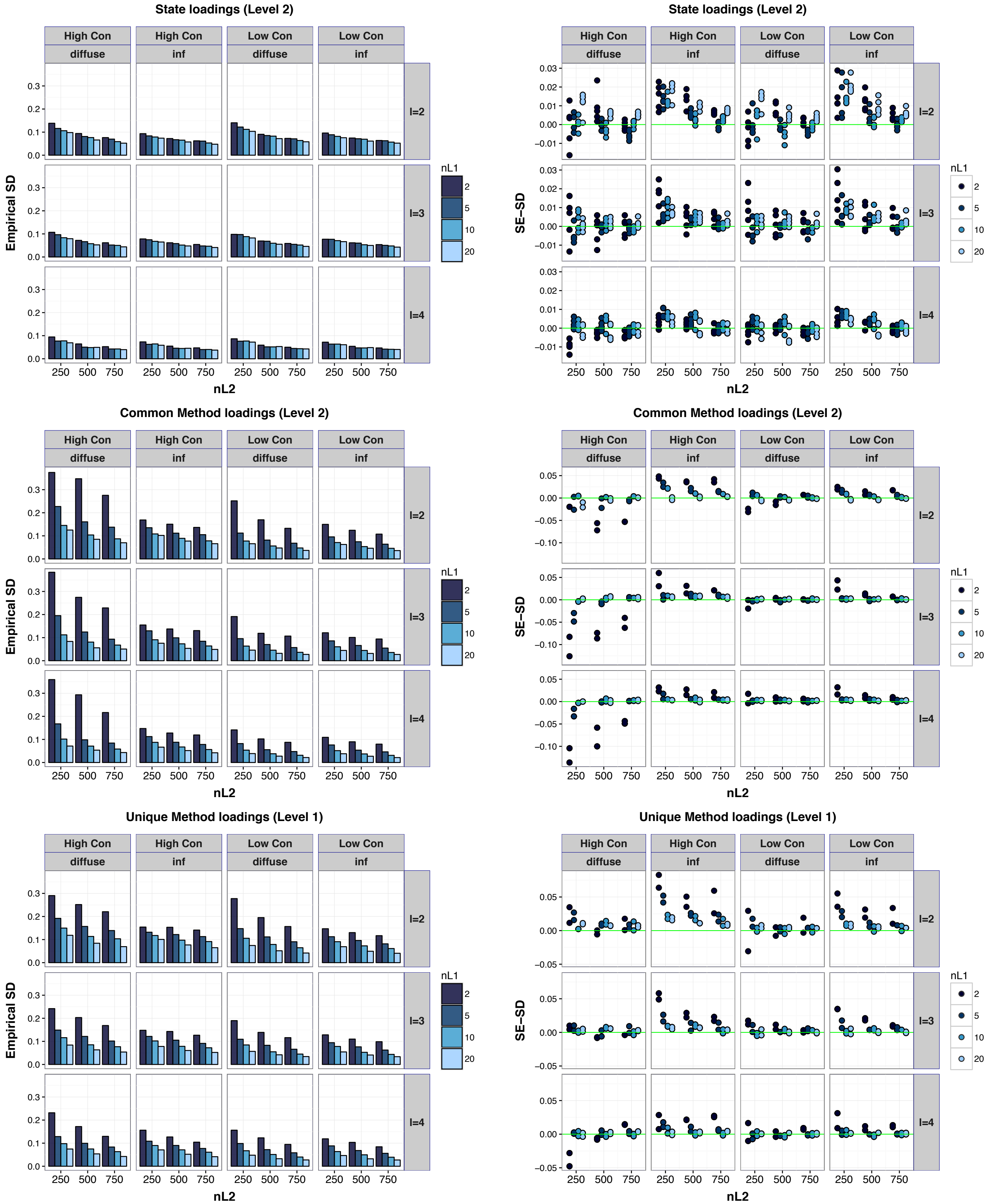

Figure A 7: Empirical SDs and standard error bias (SE - SD) for loading parameters in the LS-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

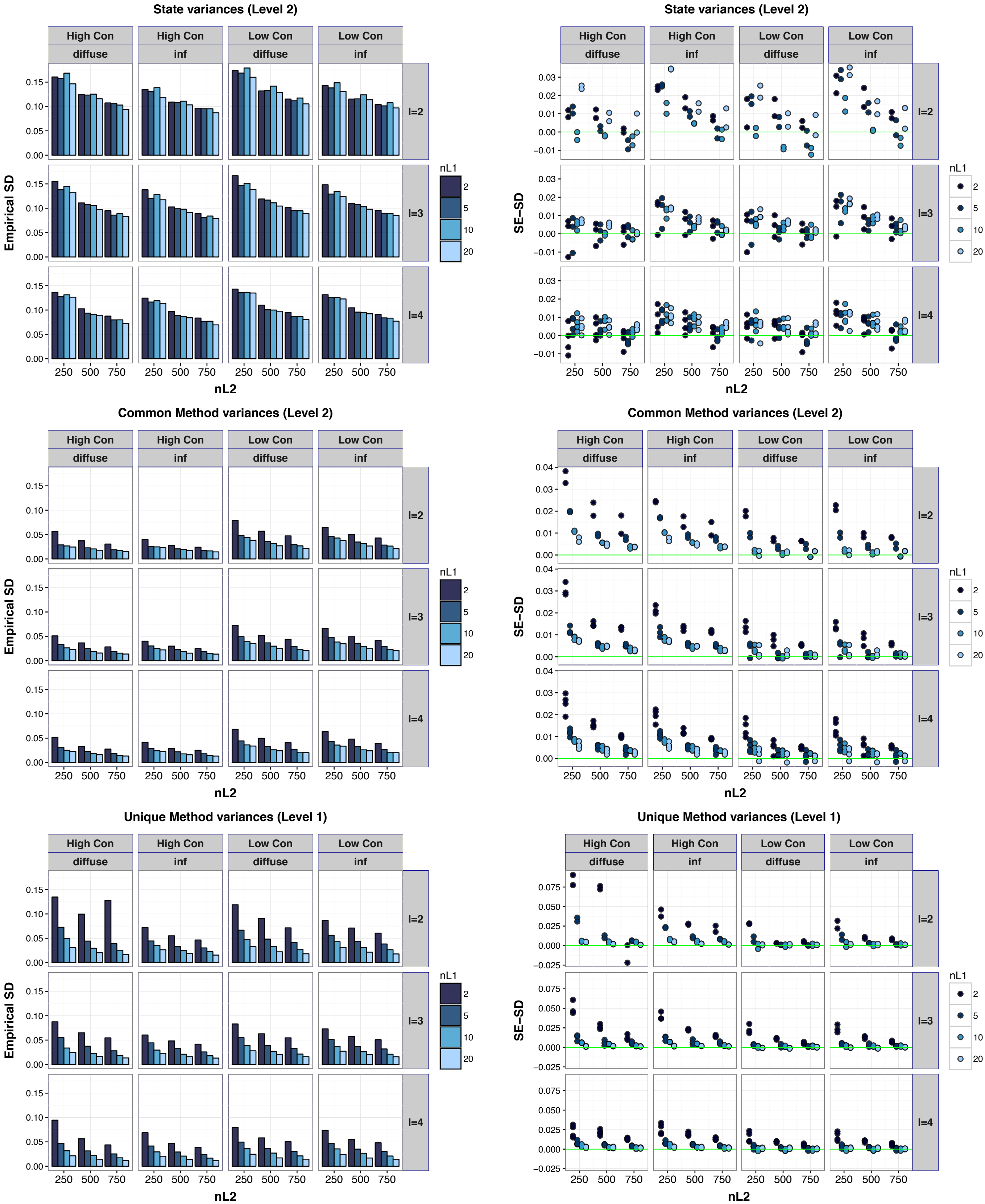

Figure A 8: Empirical SDs and standard error bias (SE - SD) for variance parameters in the LS-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

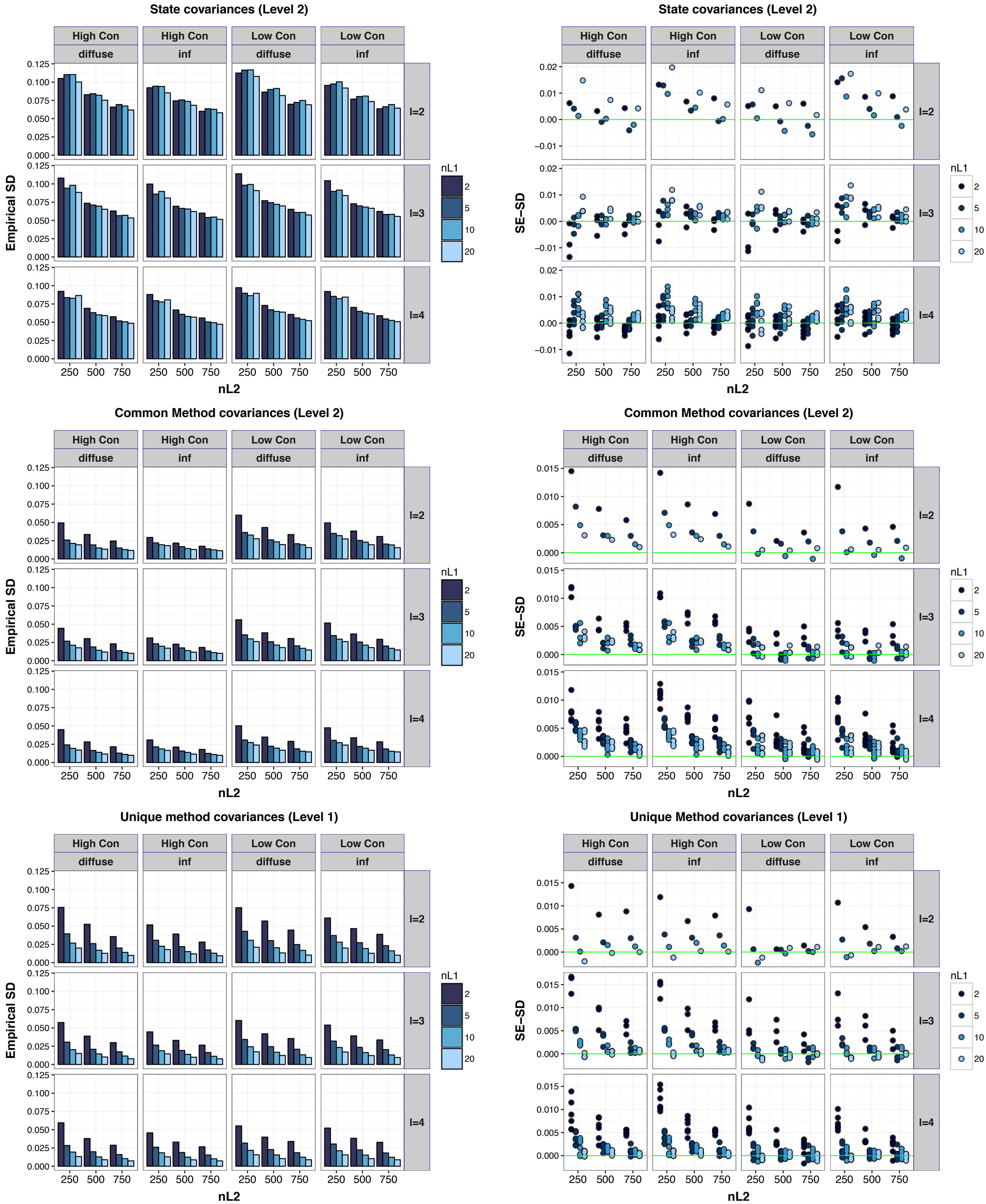

Figure A 9: Empirical SDs and standard error bias (SE - SD) for covariance parameters in the LS-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

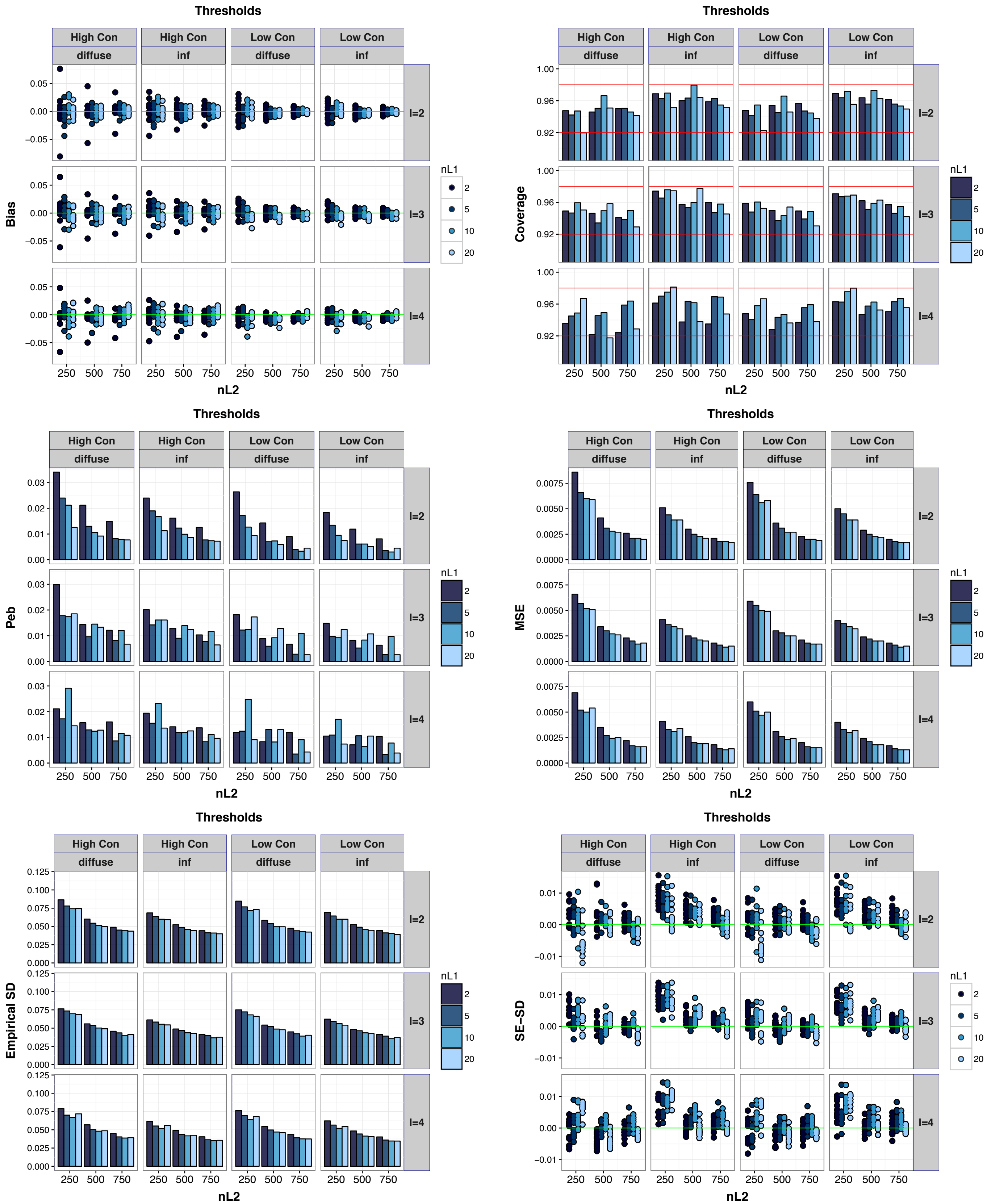

Figure A 10: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LS-Com GRM with one construct. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

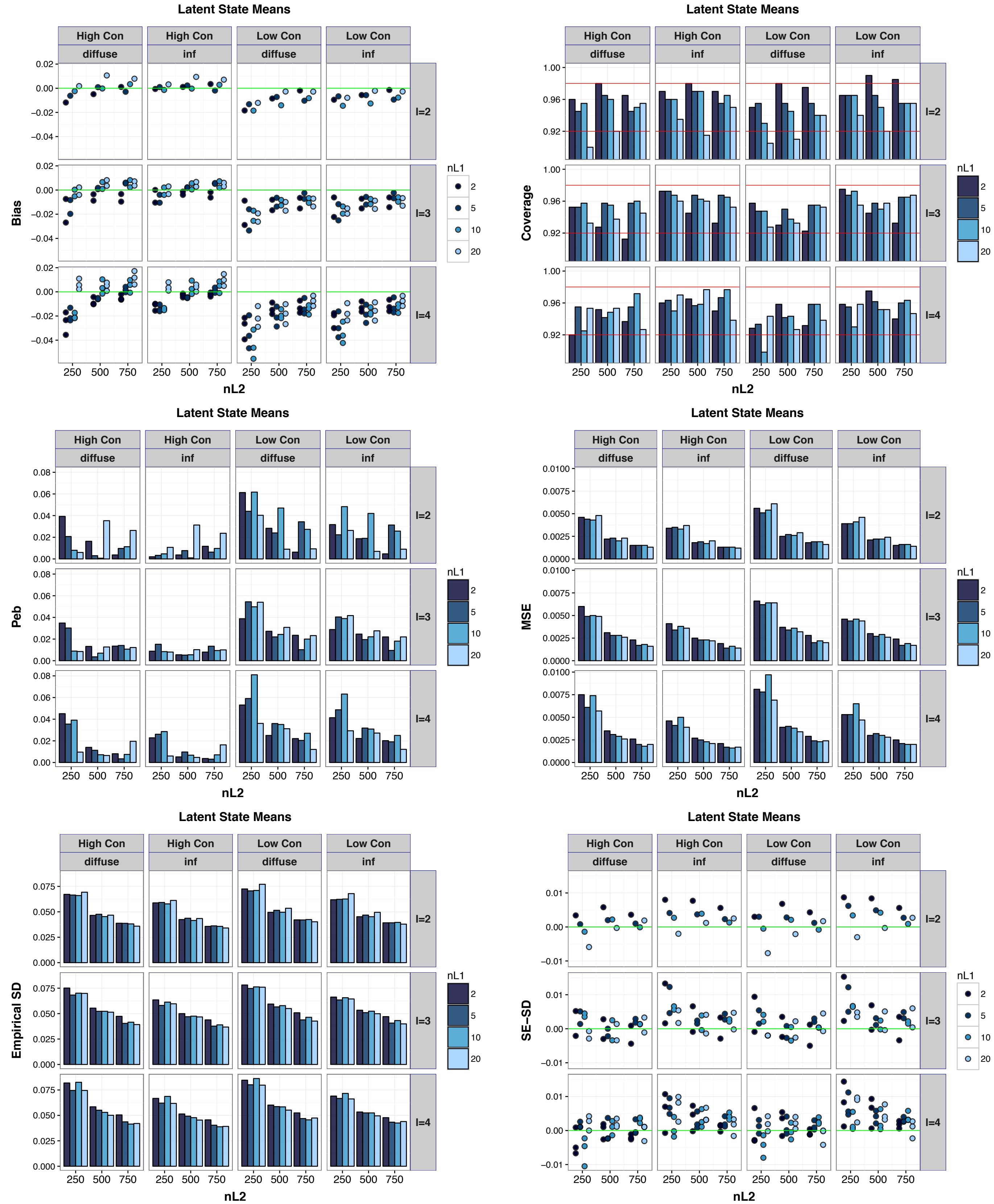

Figure A 11: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the latent state means in the LS-Com GRM with one construct. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

A.3 Simulation results LS-Com GRM. Case of two constructs ($j = 2$)

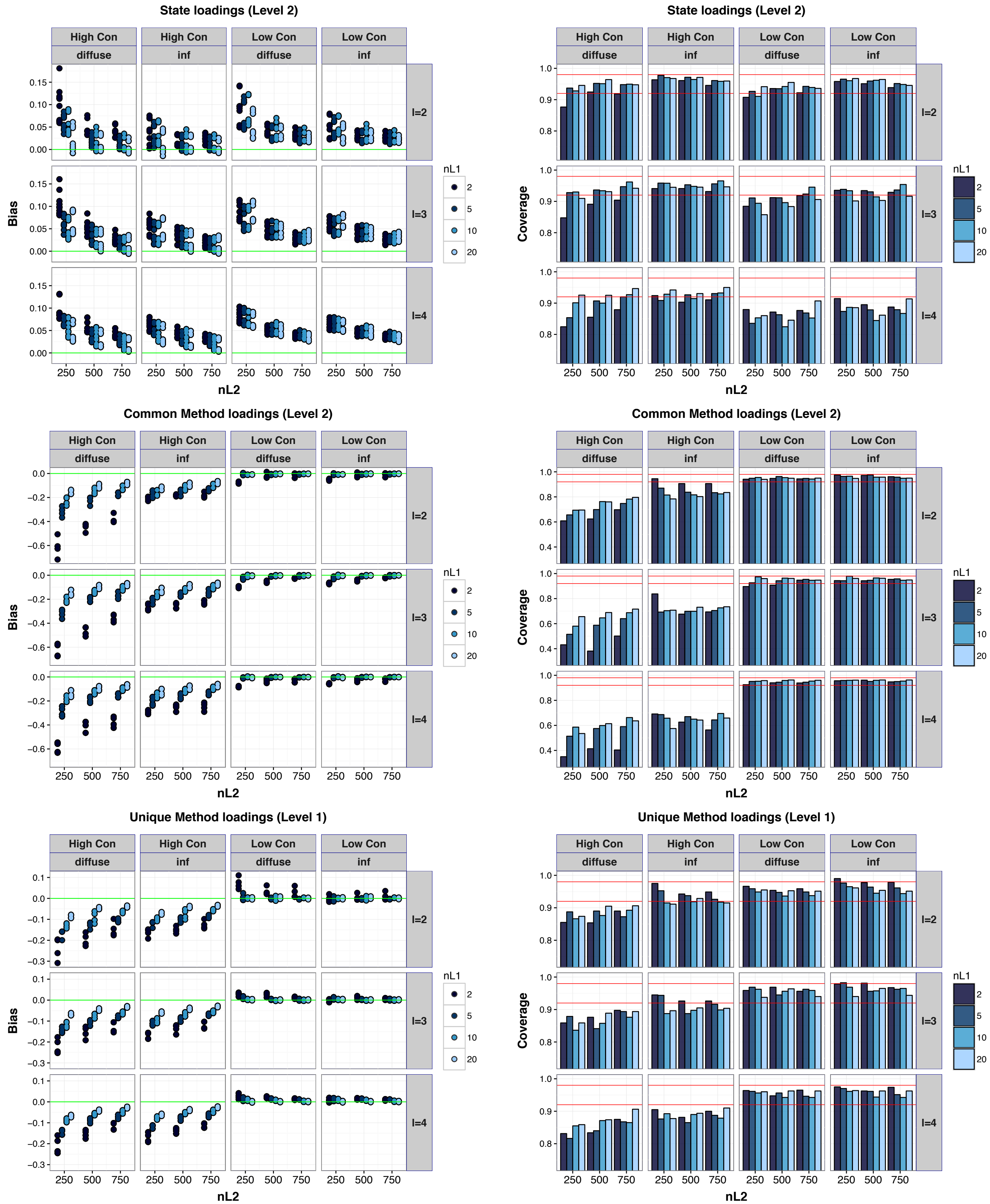

Figure A 12: Bias and 95% coverage for loading parameters in the LS-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

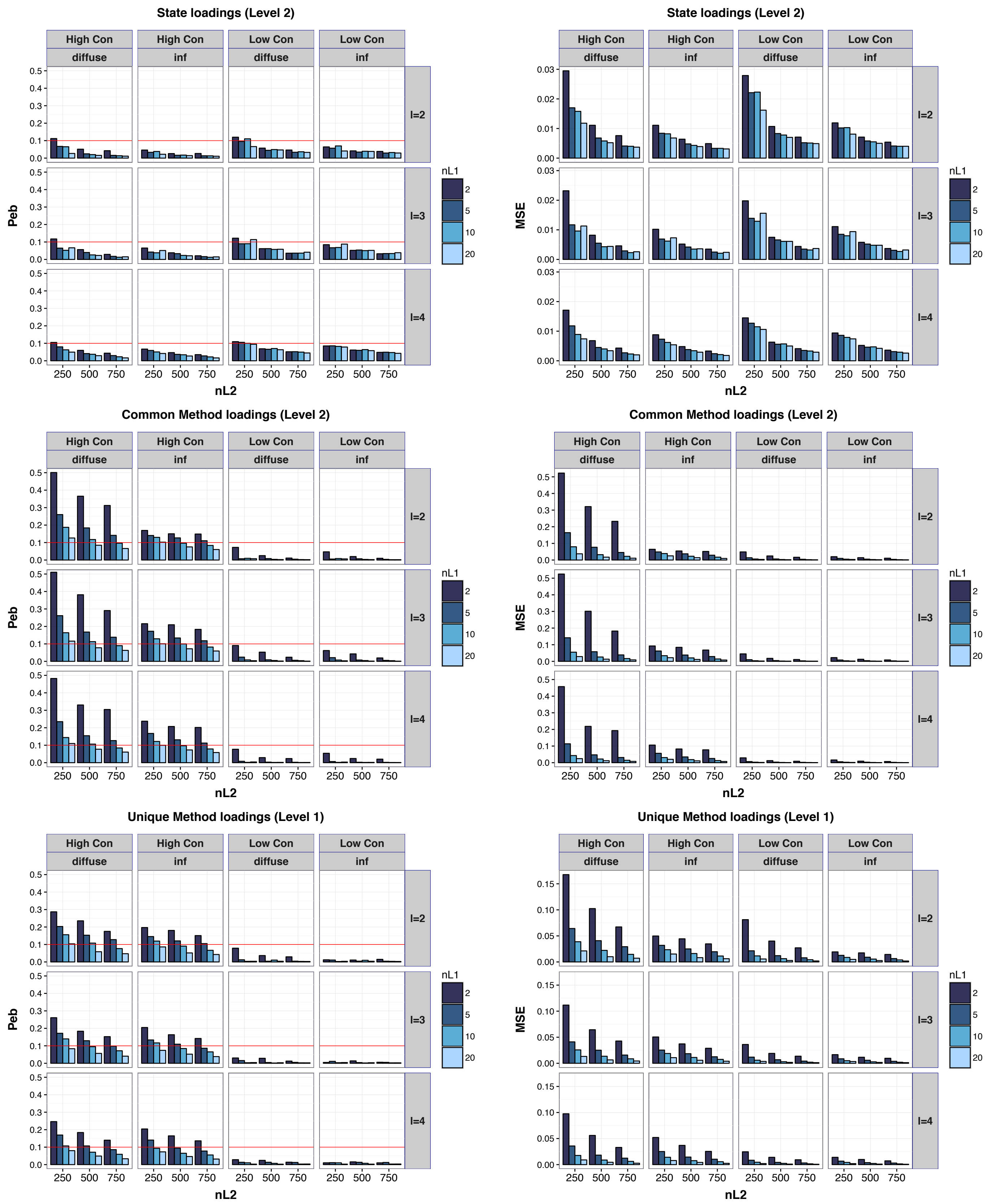

Figure A 13: Parameter estimation bias (peb) and mean squared error (MSE) for loading parameters in the LS-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

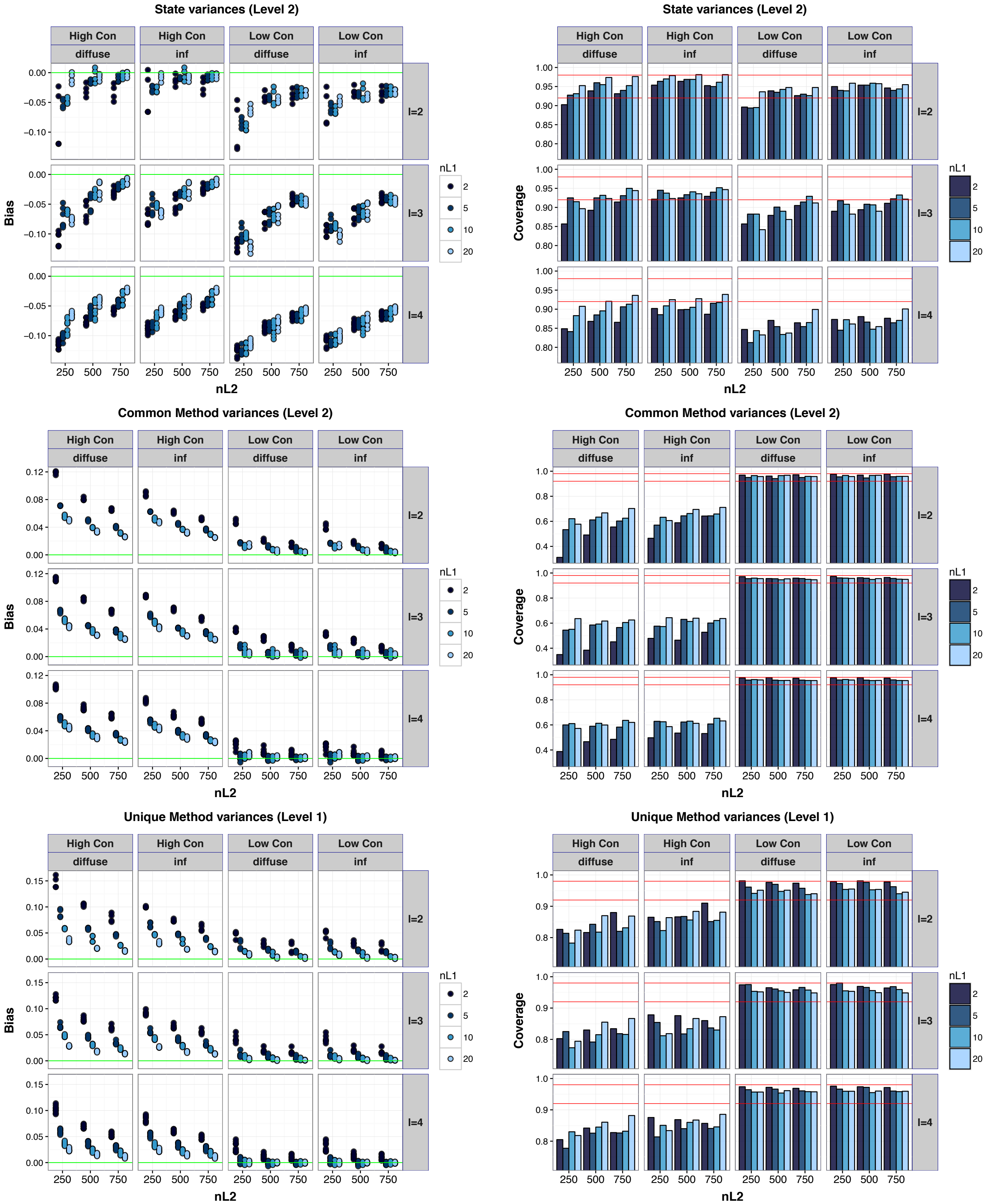

Figure A 14: Bias and 95% coverage for variance parameters in the LS-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

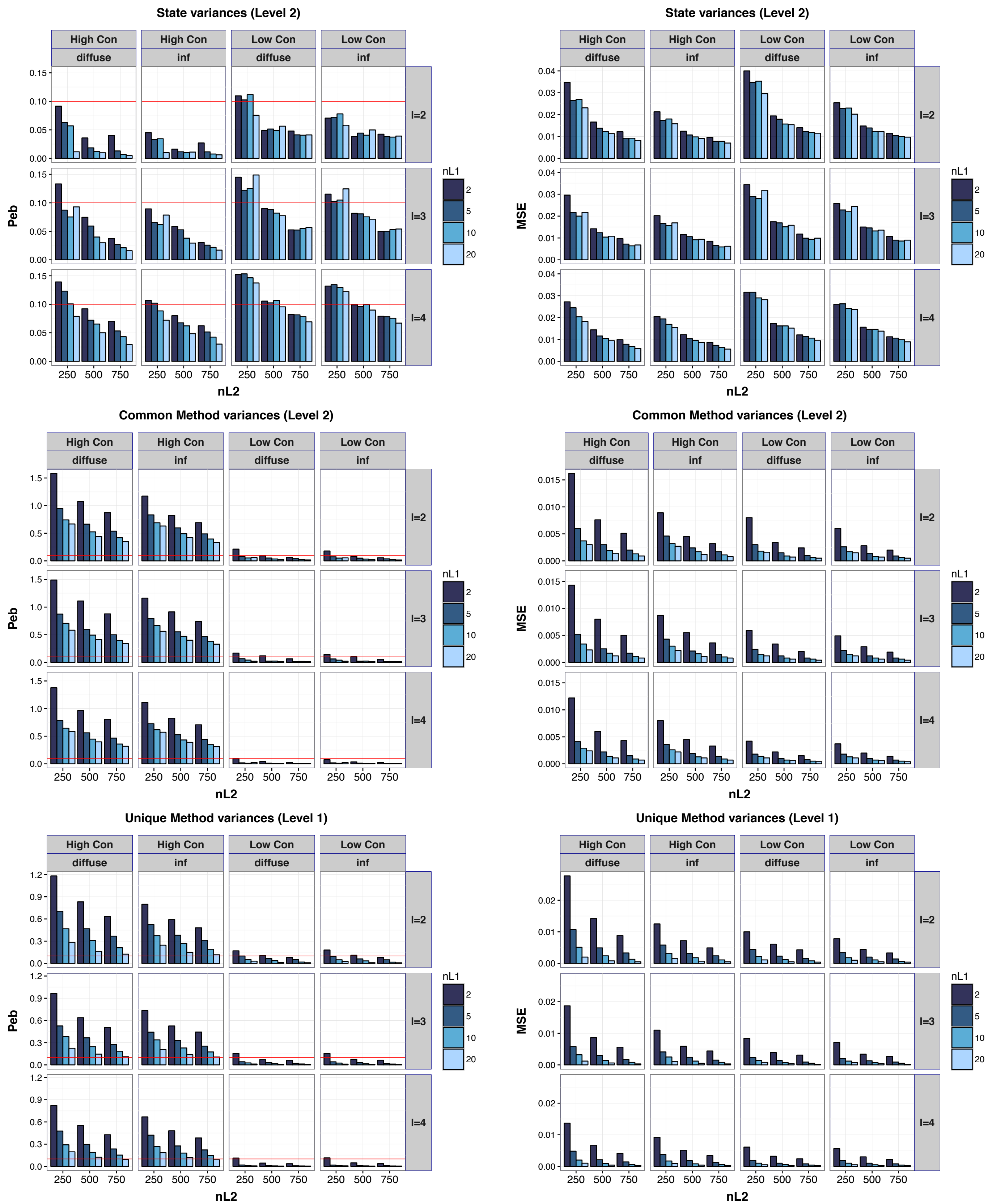

Figure A 15: Parameter estimation bias (peb) and mean squared error (MSE) for variance parameters in the LS-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

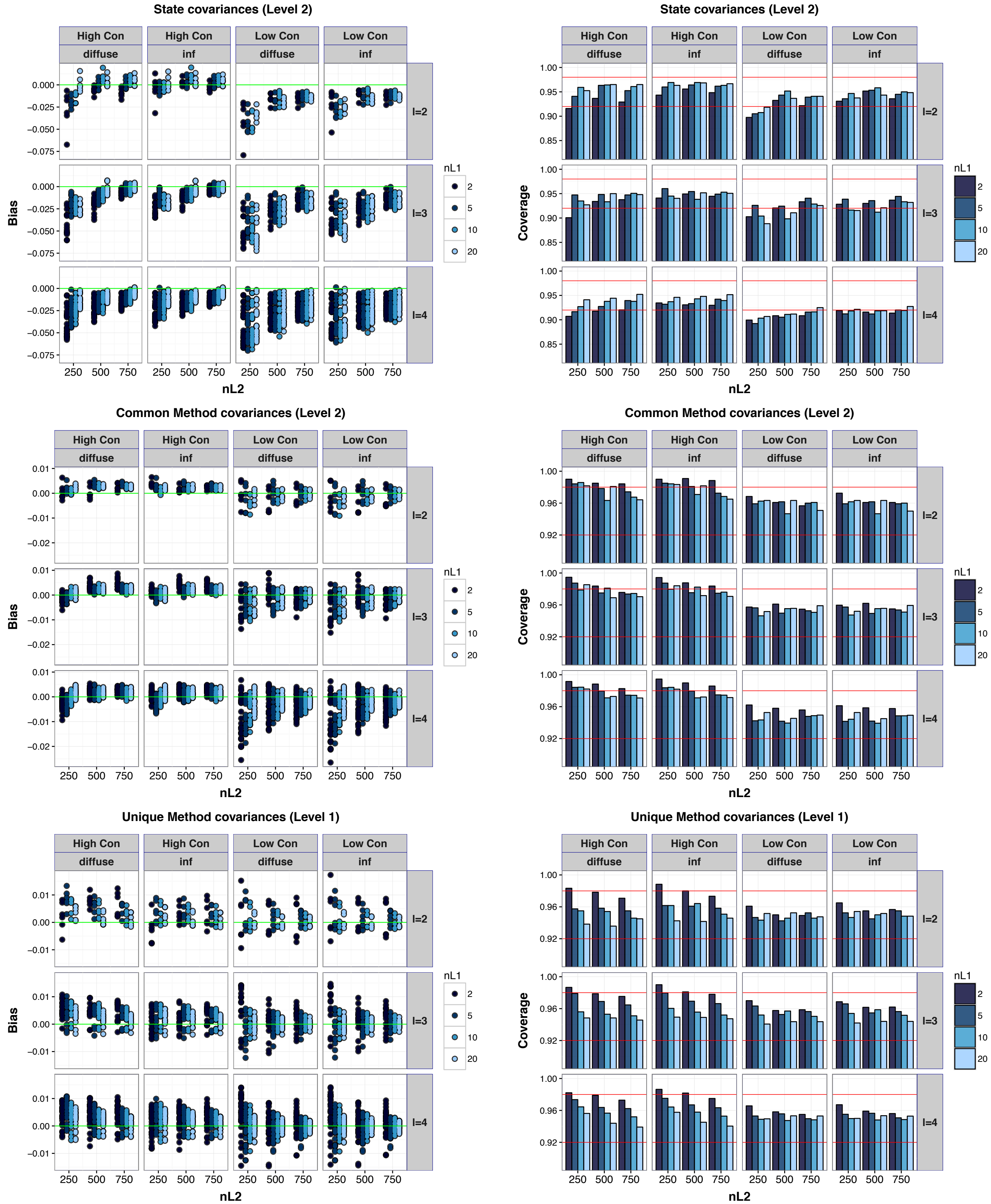

Figure A 16: Bias and 95% coverage for covariance parameters in the LS-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

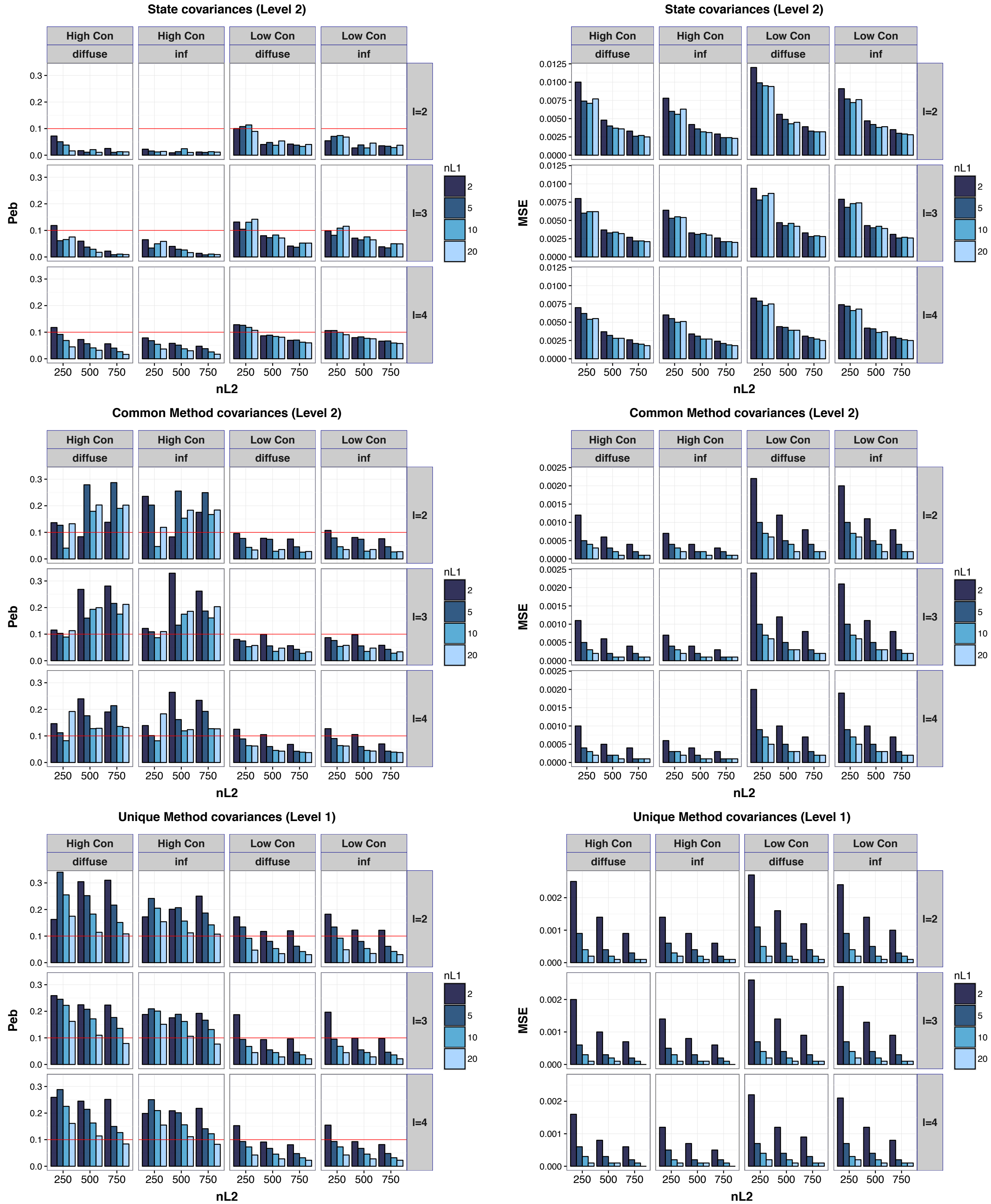

Figure A 17: Parameter estimation bias (peb) and mean squared error (MSE) for covariance parameters in the LS-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

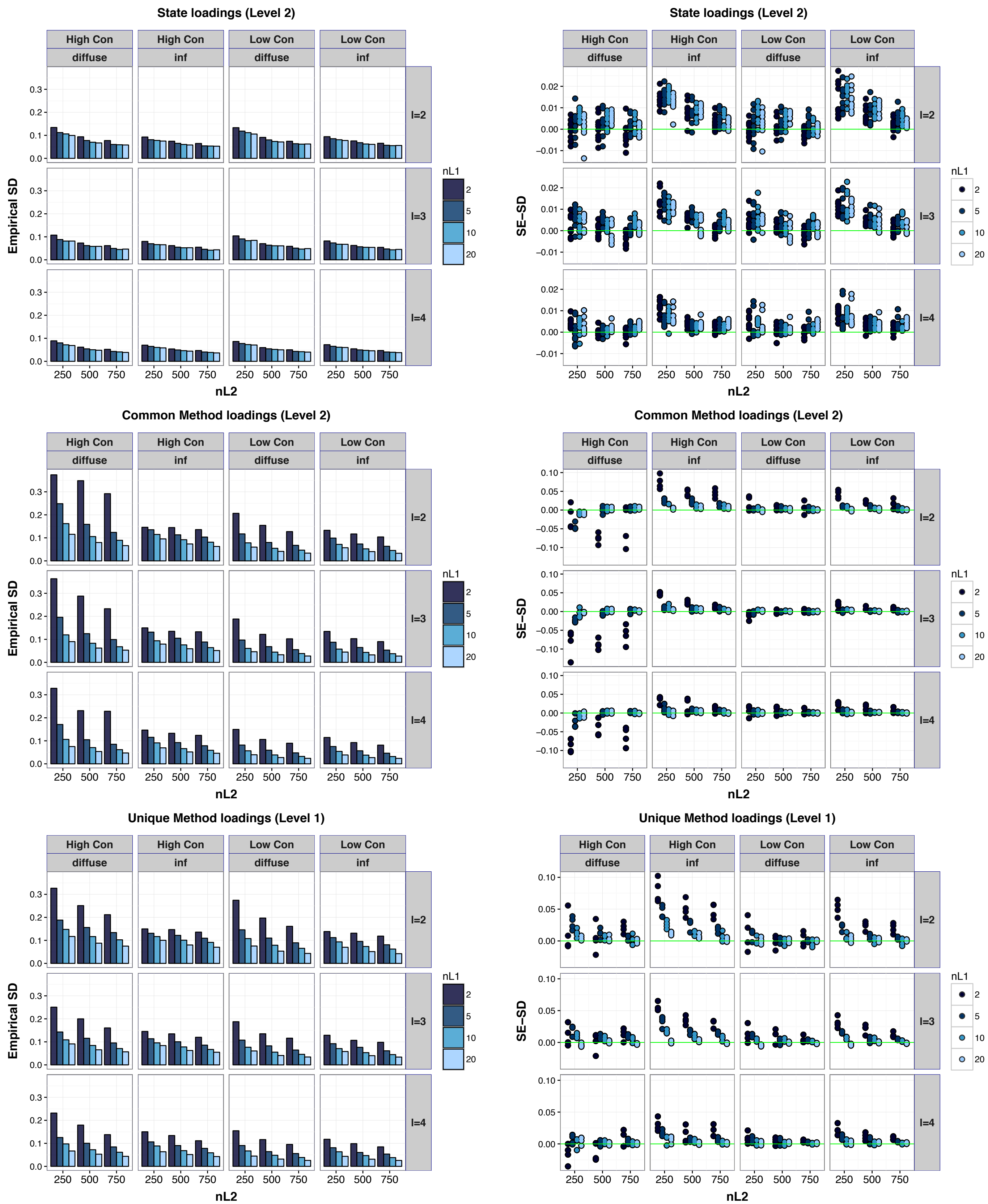

Figure A 18: Empirical SDs and standard error bias (SE - SD) for loading parameters in the LS-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

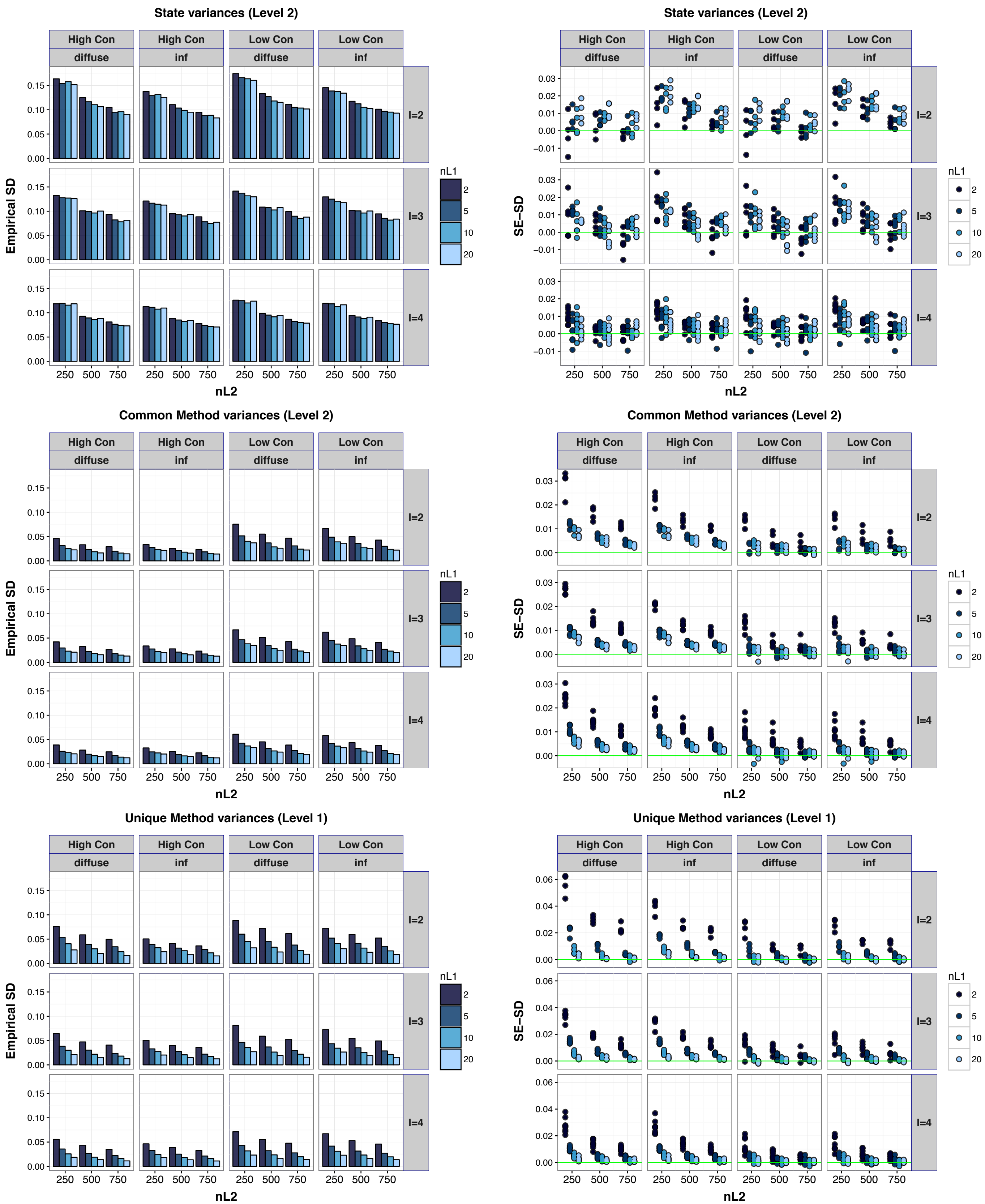

Figure A 19: Empirical SDs and standard error bias (SE - SD) for variance parameters in the LS-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

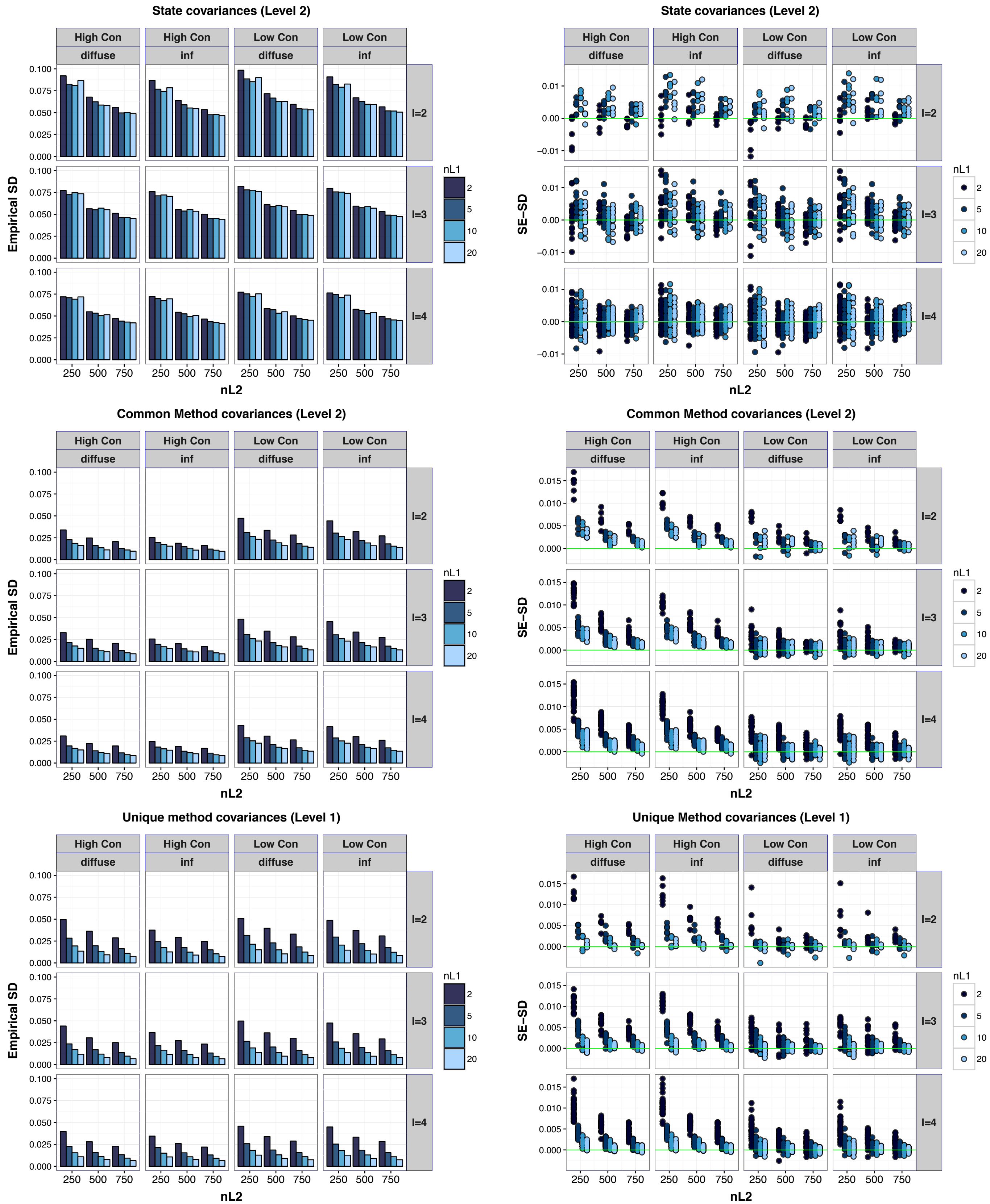

Figure A 20: Empirical SDs and standard error bias (SE - SD) for covariance parameters in the LS-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

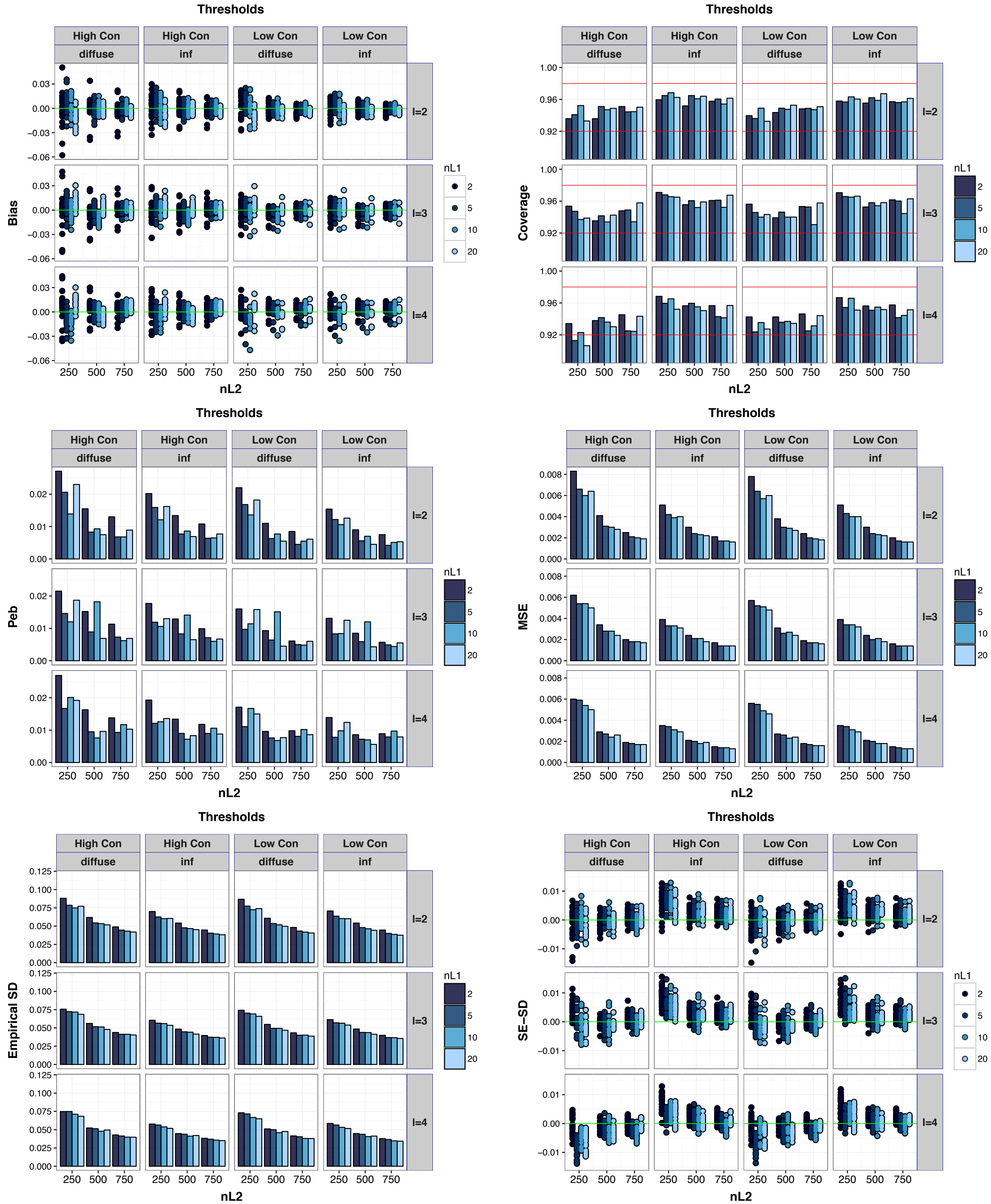

Figure A 21: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LS-Com GRM with two constructs. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

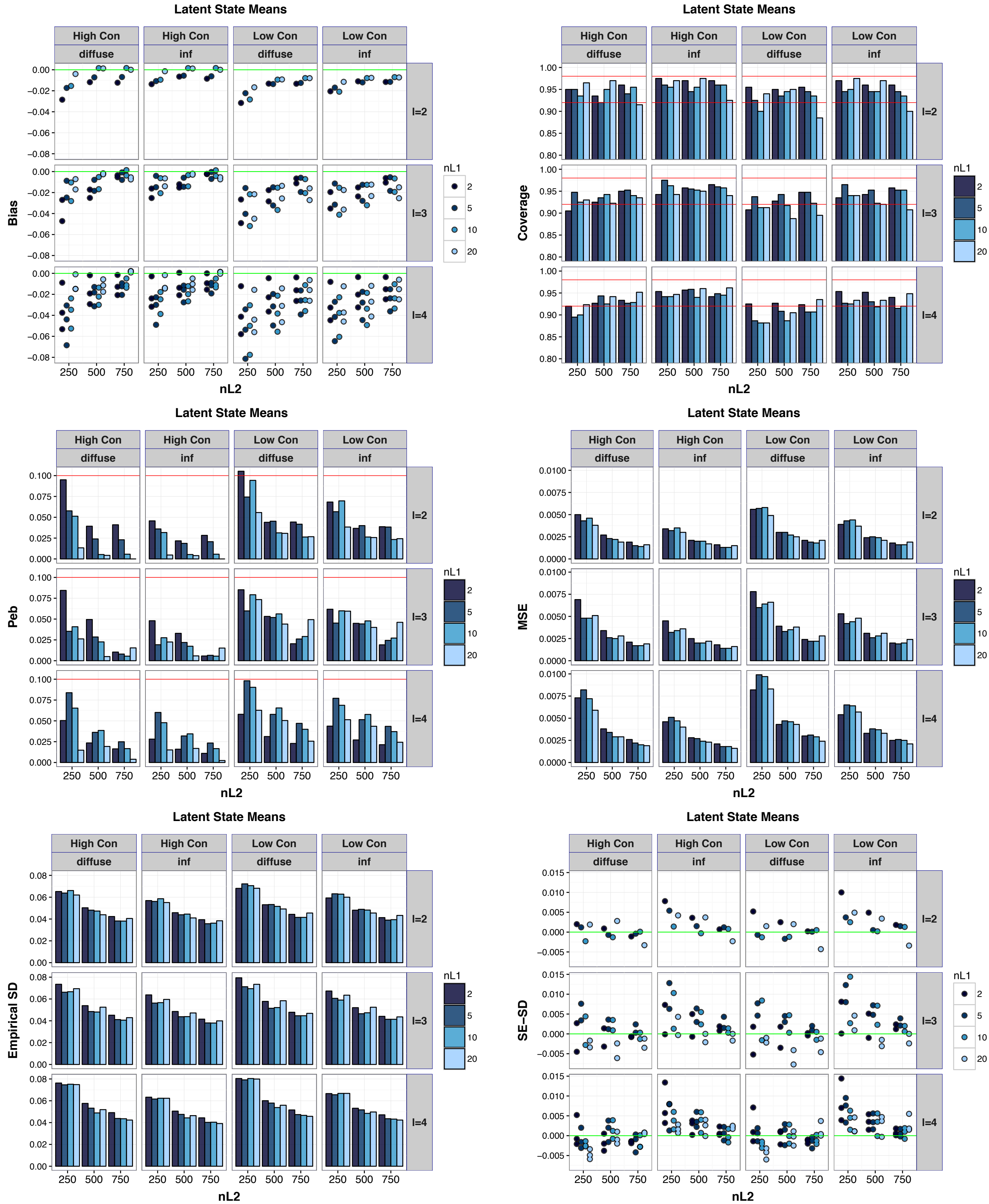

Figure A 22: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the latent state means in the LS-Com GRM with two constructs. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

B Monte Carlo simulation study LST-Com GRM

B.1 Population parameters in the LST-Com GRM simulation

Table B 1: Population values in the LST-Com GRM Monte Carlo simulation study

Population parameters in the LST-Com GRM simulation				
	Low Con		High Con	
Parameter	$j = 1$	$j = 2$	$j = 1$	$j = 2$
Within-level				
UMS				
Loadings	1	1	1	1
Variances	0.16	0.16	0.065	0.065
Covariances	-	0.024	-	0.006
UMT				
Variances	0.16	0.16	0.065	0.065
Covariances	0.096	0.048	0.024	0.012
Between-level				
Traits				
Loadings	0.5	0.5	0.7857	0.7857
Variances	0.49	0.49	0.49	0.49
Covariances	0.392	0.196	0.392	0.196
State Residuals				
Loadings	1 / 0.5833	1 / 0.5833	1 / 0.9167	1 / 0.9167
Variances	0.36	0.36	0.36	0.36
Covariances	-	0.072	-	0.072
CMS				
Loadings	1	1	1	1
Variances	0.1225	0.1225	0.065	0.065
Covariances	-	0.0184	-	0.0098
CMT				
Variances	0.1225	0.1225	0.065	0.065
Covariances	0.0735	0.0368	0.039	0.0195

Note. Values of the population parameters for all parameters of the same parameter type in the respective condition. Population values did not vary between the parameters of one parameter class. Note that covariance parameters reported for the multi-construct conditions ($j = 2$) correspond to the covariances of the respective factors between constructs. Covariances within one construct are identical in the mono-construct and multi-construct conditions and reported under $j = 1$. *CMS*: common method state residual variables; *CMT*: common method trait variables; High Con; high consistency condition; j: number of constructs; Low Con: Low consistency condition; *UMS*: unique method state residual variables; *UMT*: unique method trait variables.

B.2 Simulation results LST-Com GRM. Case of one construct ($j = 1$)

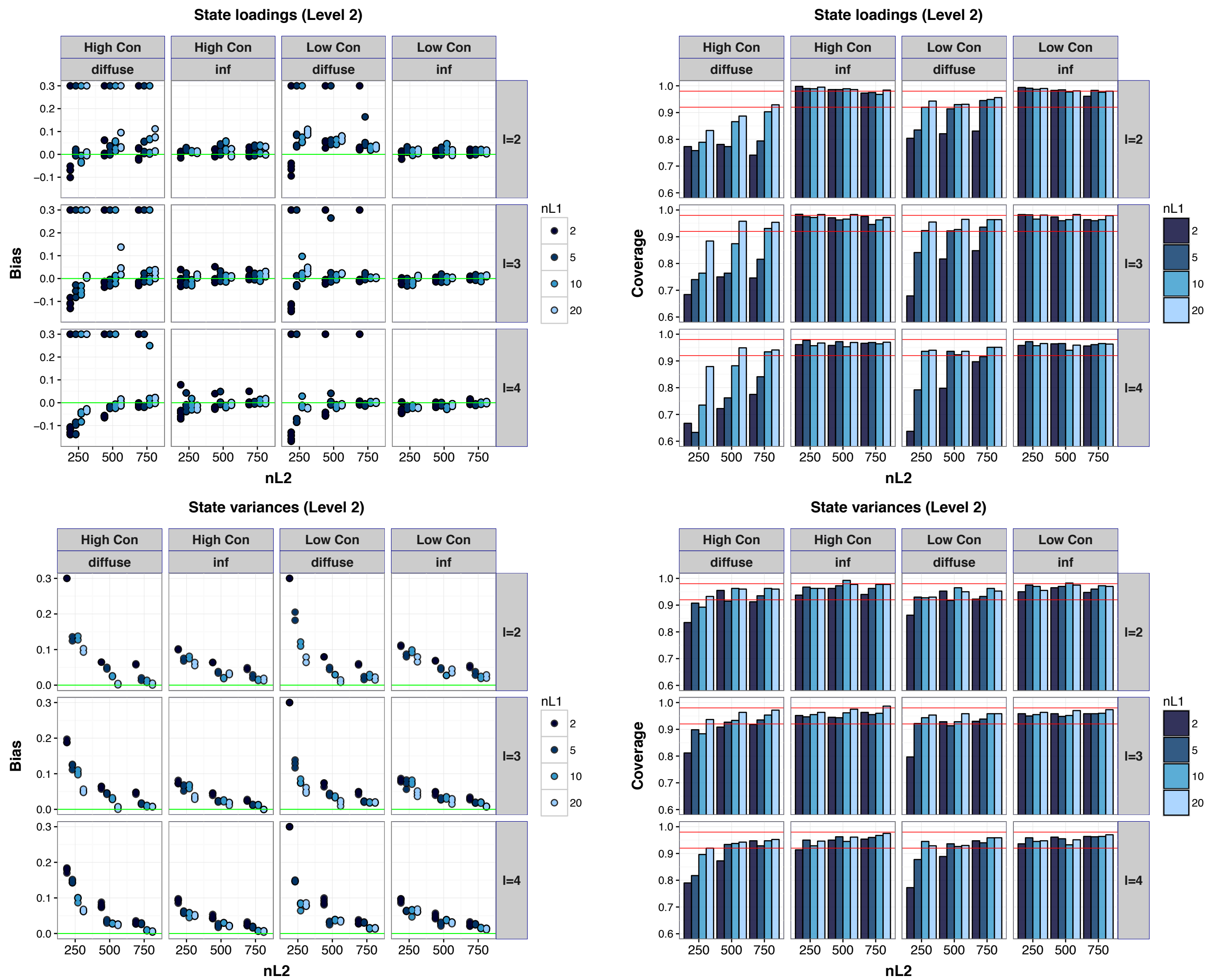

Figure B 1: Bias and 95% coverage for latent state residual factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and $<$ -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

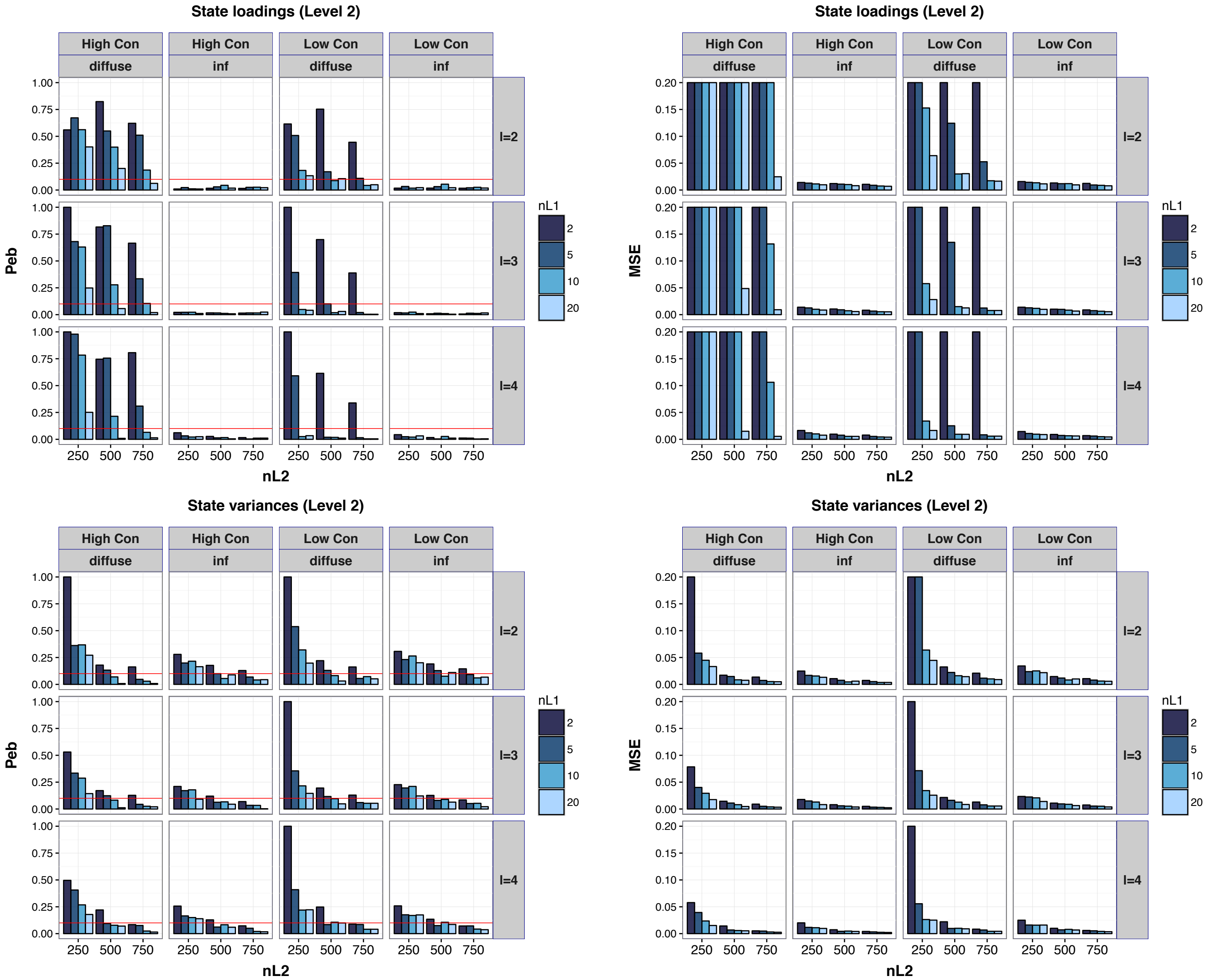

Figure B 2: Parameter estimation bias (peb) and mean squared error (MSE) for latent state residual factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

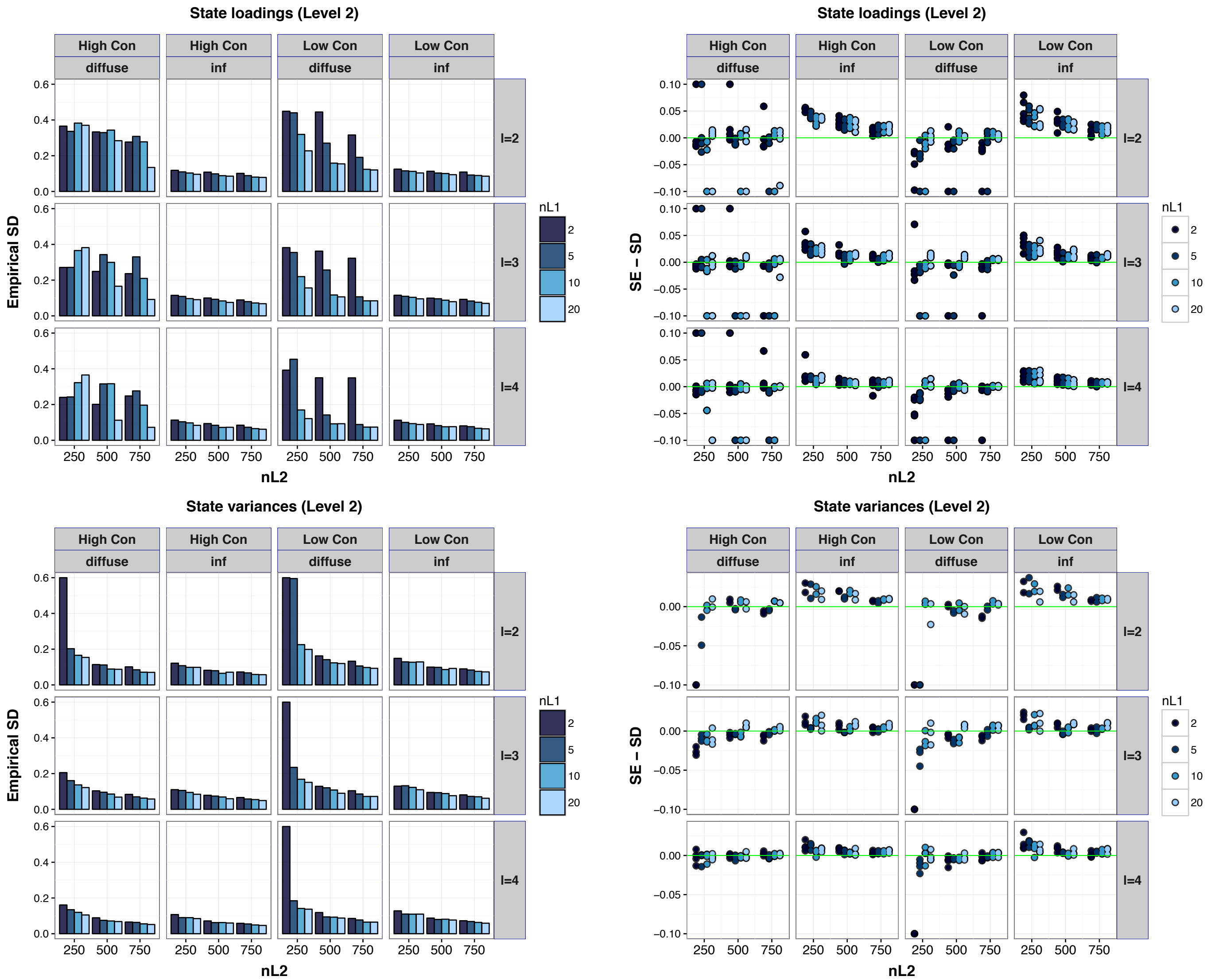

Figure B 3: Empirical SDs and standard error bias (SE - SD) for latent state residual factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

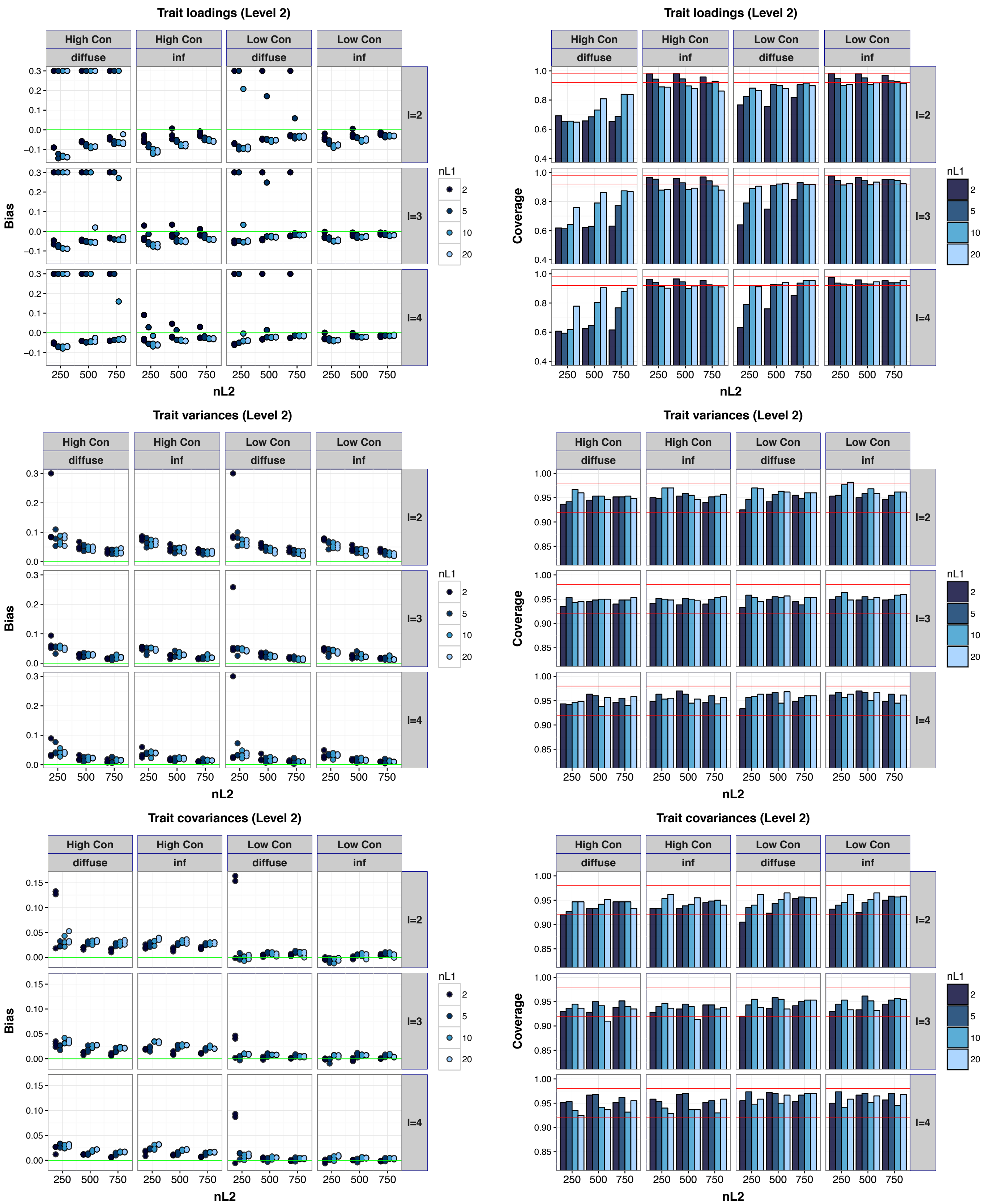

Figure B 4: Bias and 95% coverage for latent trait factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values> 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

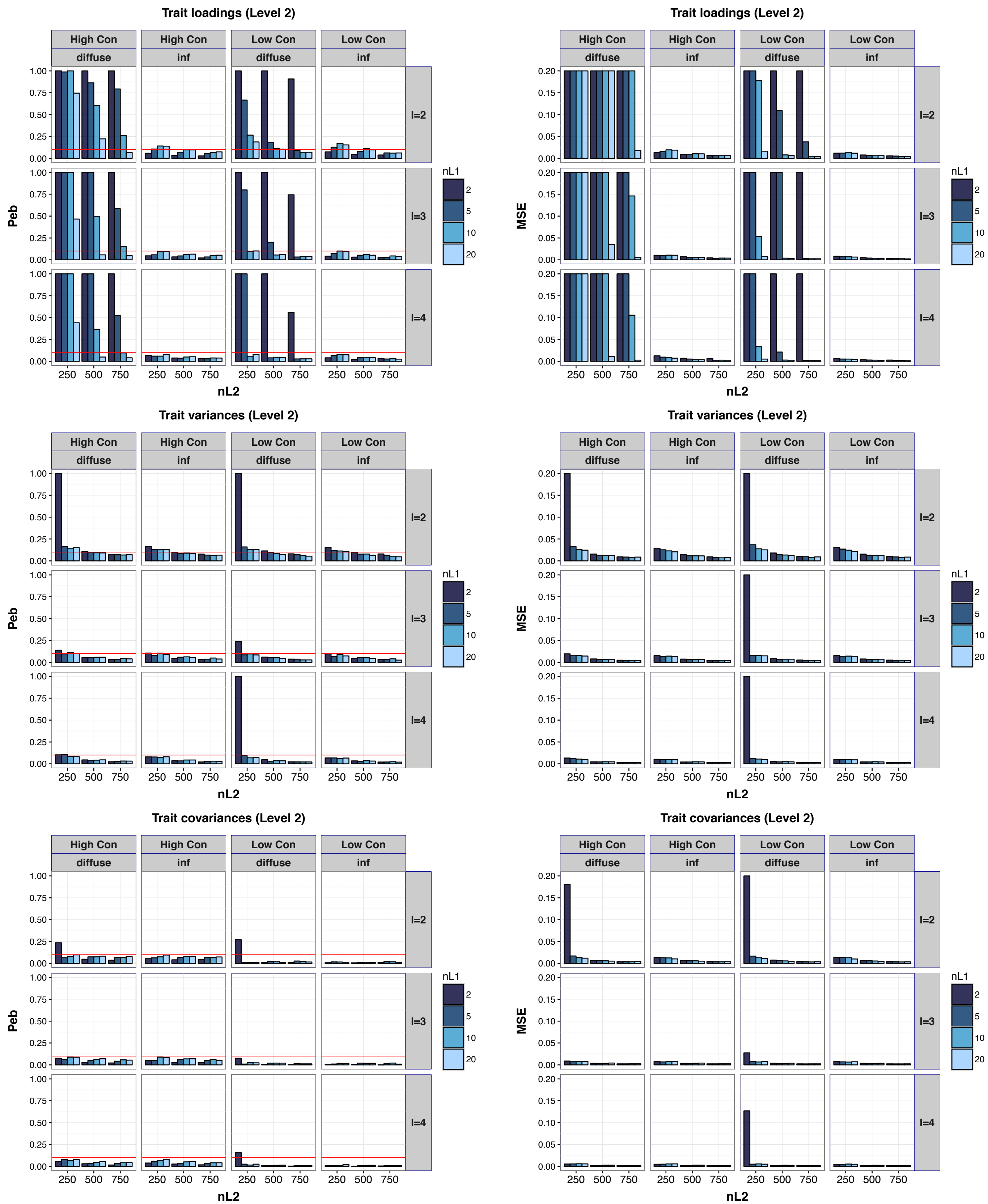

Figure B 5: Parameter estimation bias (peb) and mean squared error (MSE) for latent trait factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

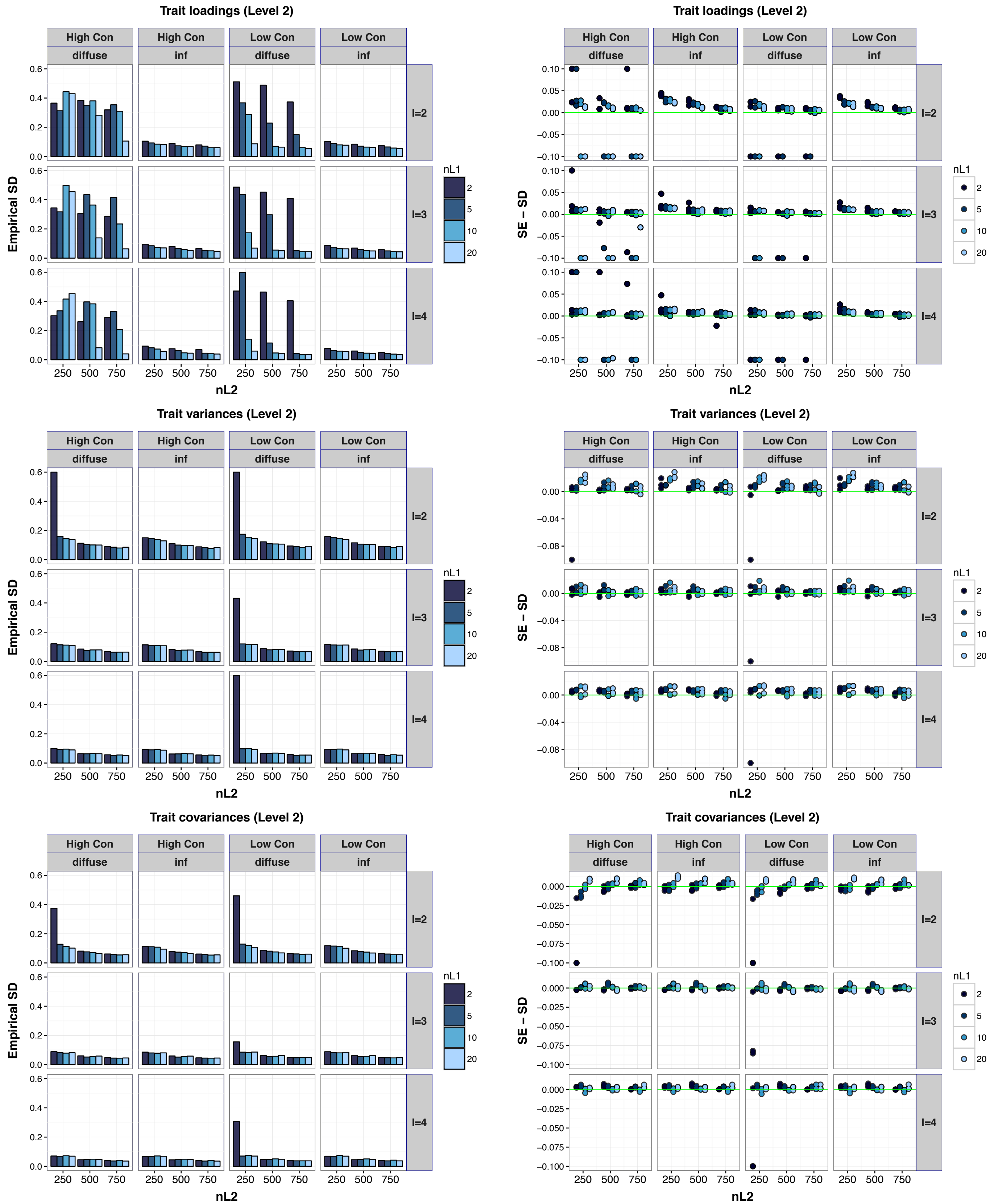

Figure B 6: Empirical SDs and standard error bias (SE - SD) for latent trait factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

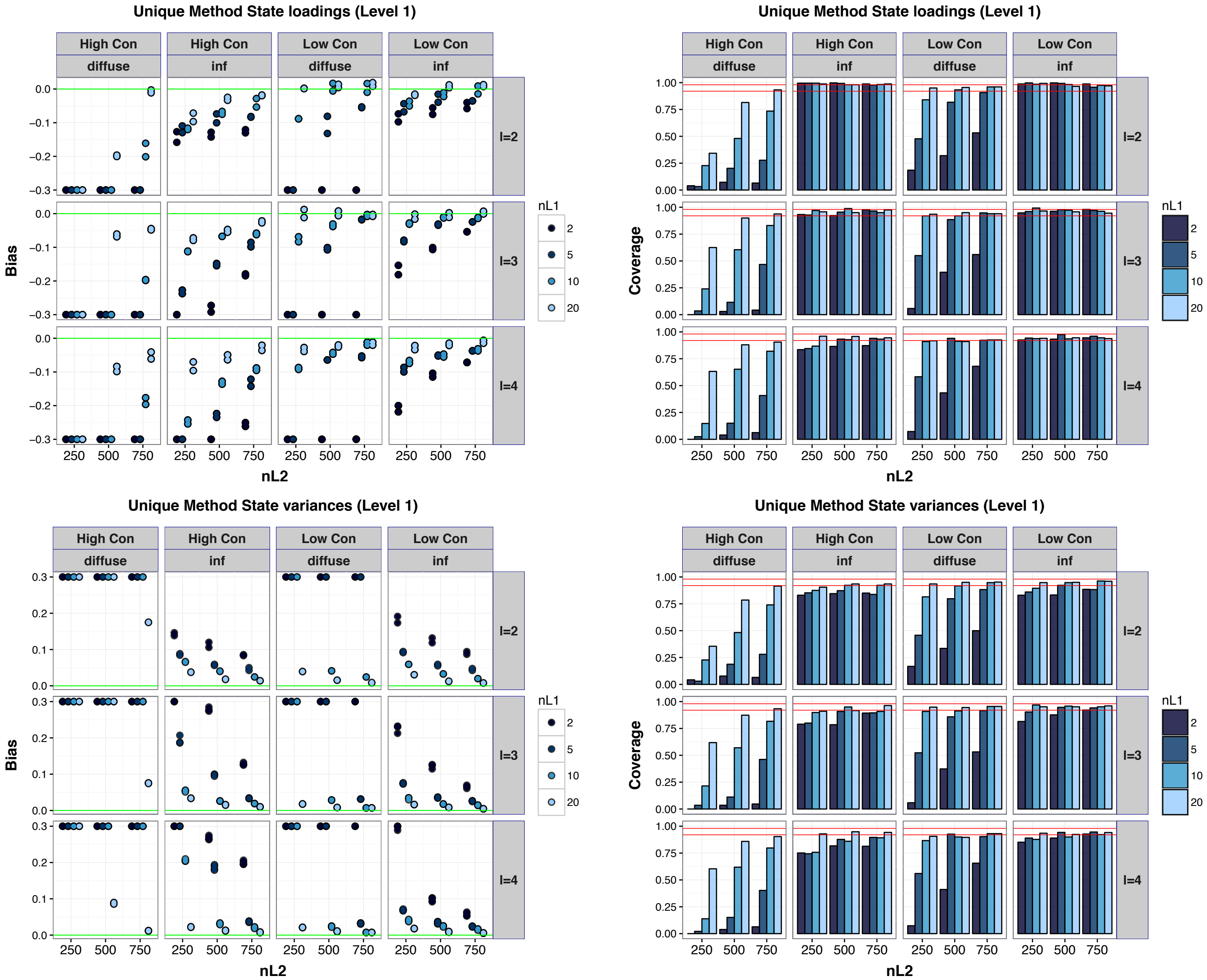

Figure B 7: Bias and 95% coverage for unique method state residual factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

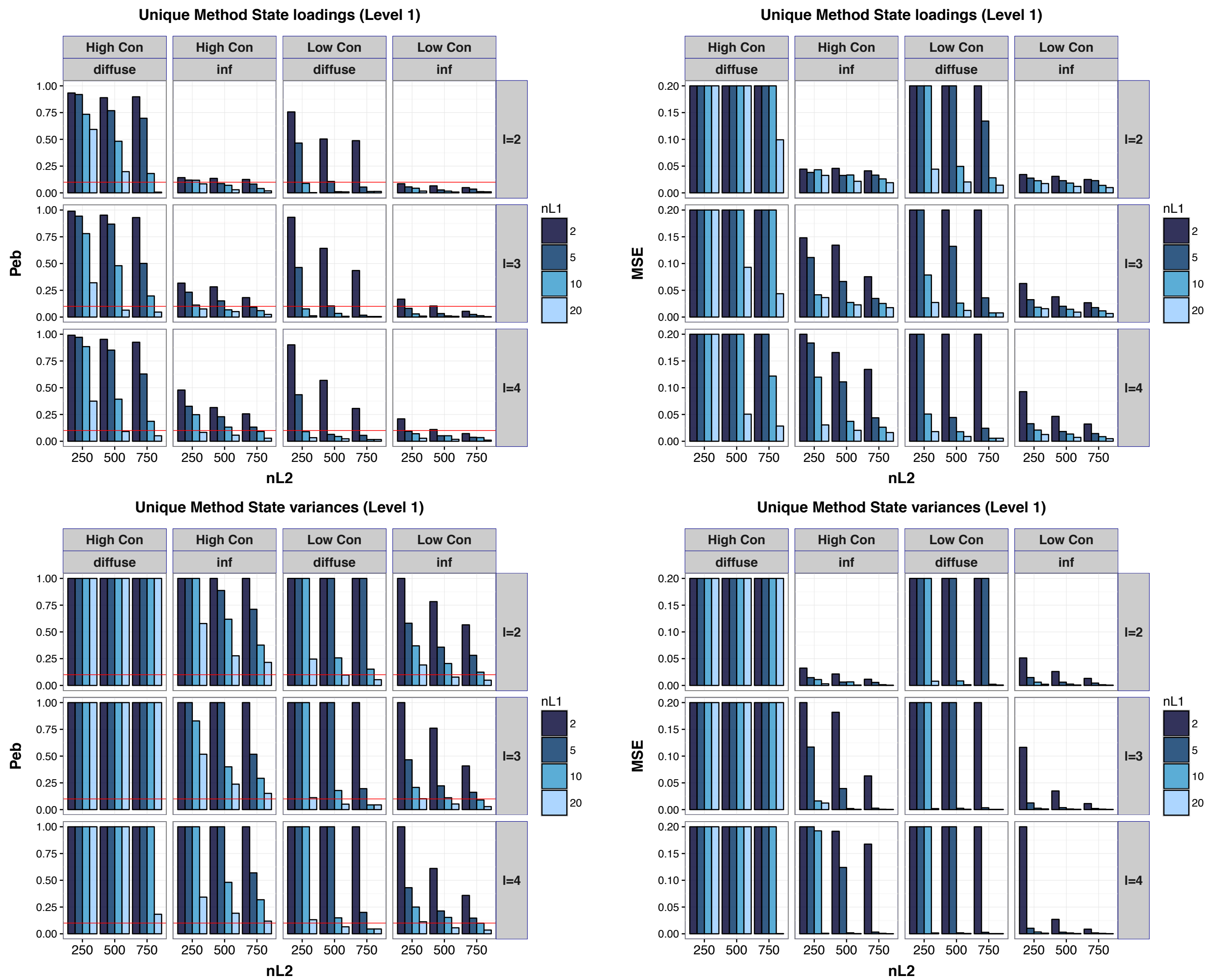

Figure B 8: Parameter estimation bias (peb) and mean squared error (MSE) for unique method state residual factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

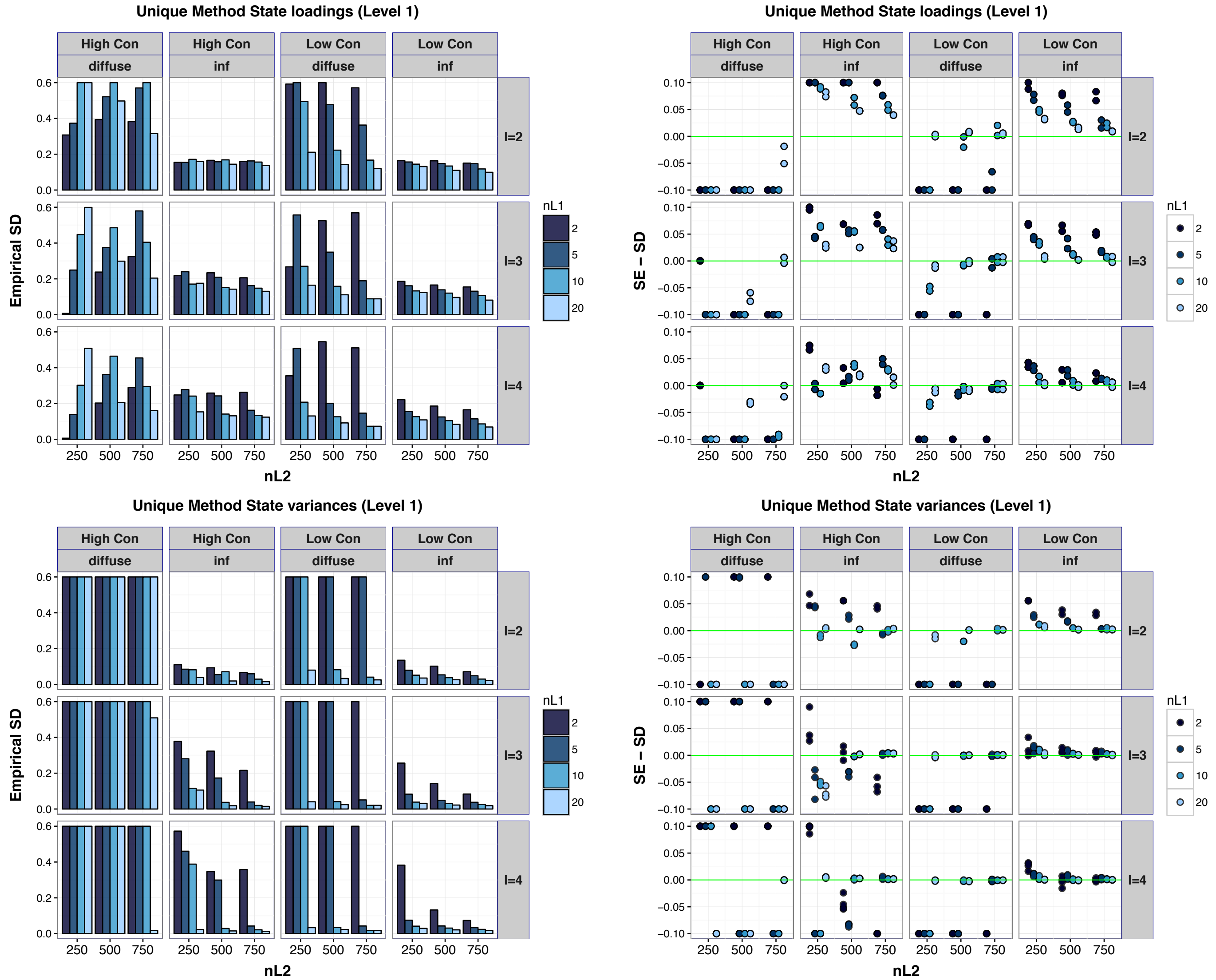

Figure B 9: Empirical SDs and standard error bias (SE - SD) for unique method state residual factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

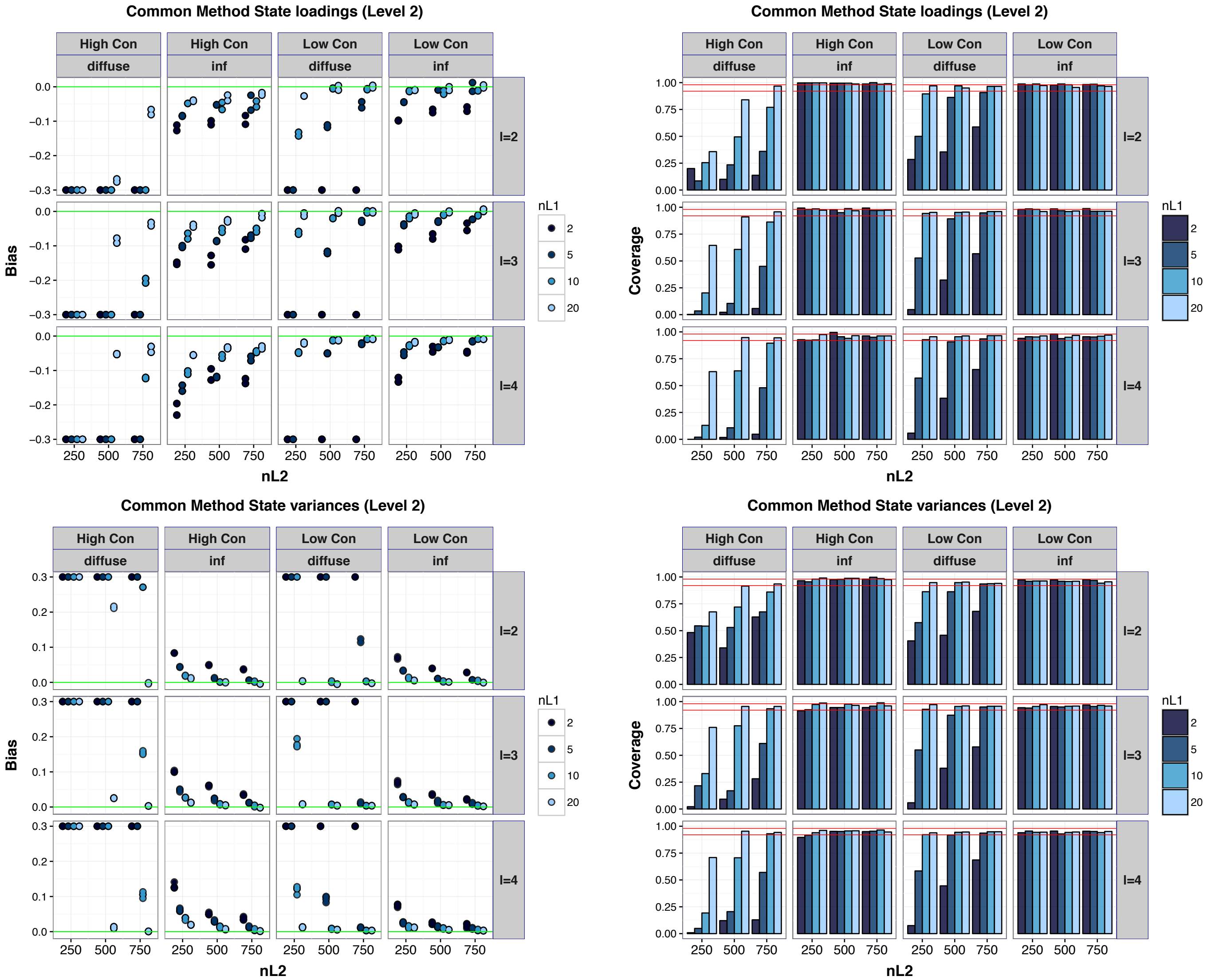

Figure B 10: Bias and 95% coverage for common method state residual factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

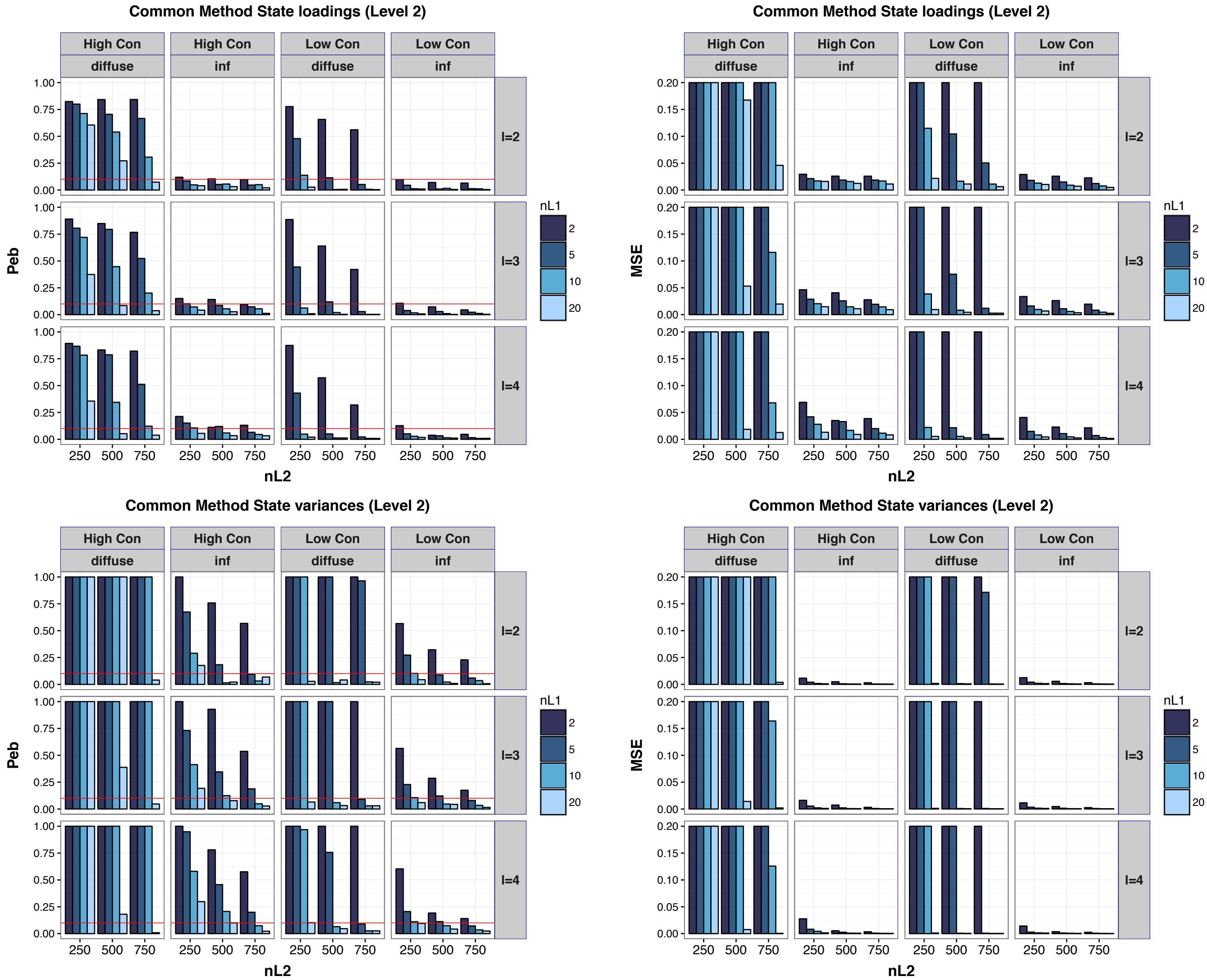

Figure B 11: Parameter estimation bias (peb) and mean squared error (MSE) for common method state residual factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

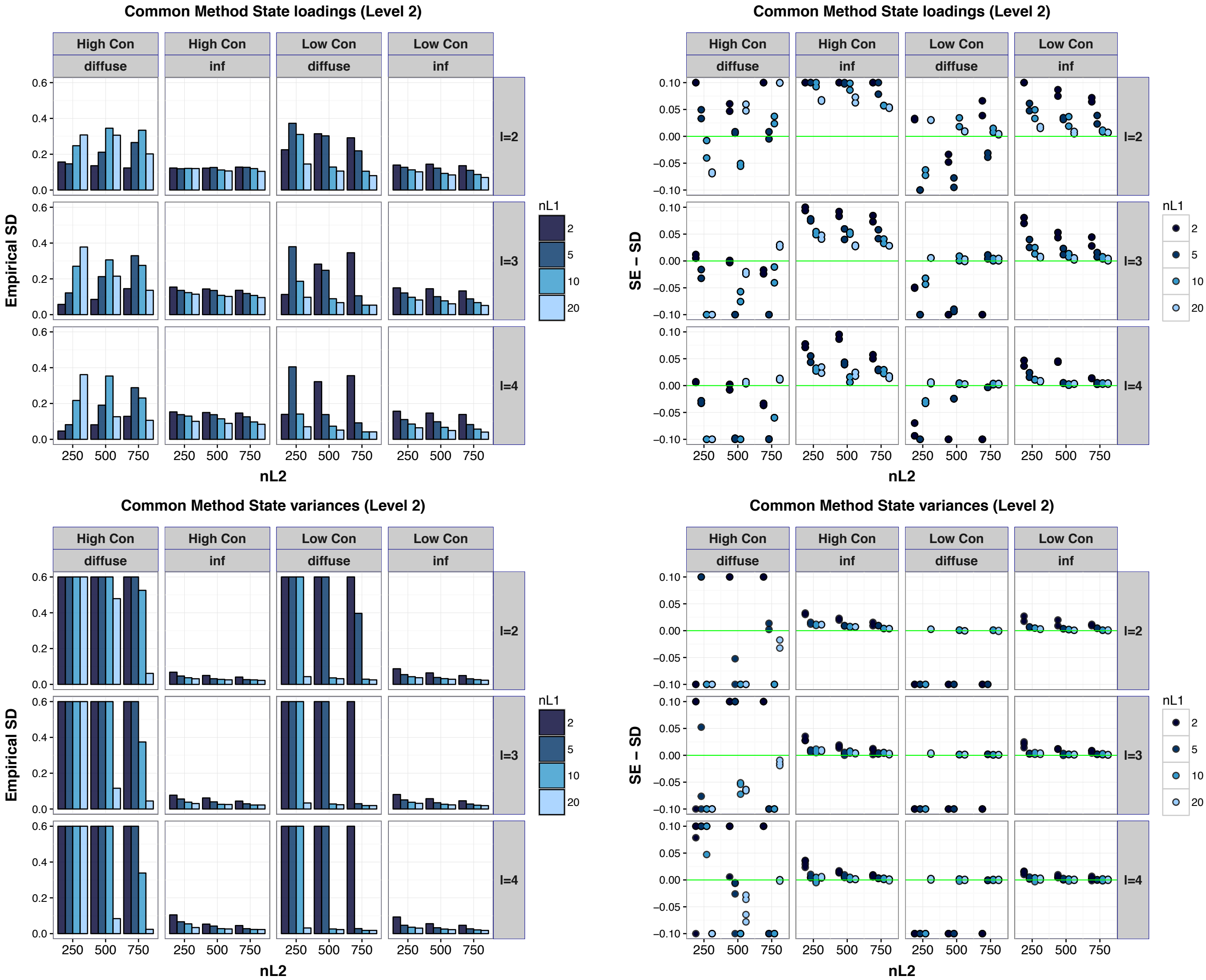

Figure B 12: Empirical SDs and standard error bias (SE - SD) for common method state residual factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

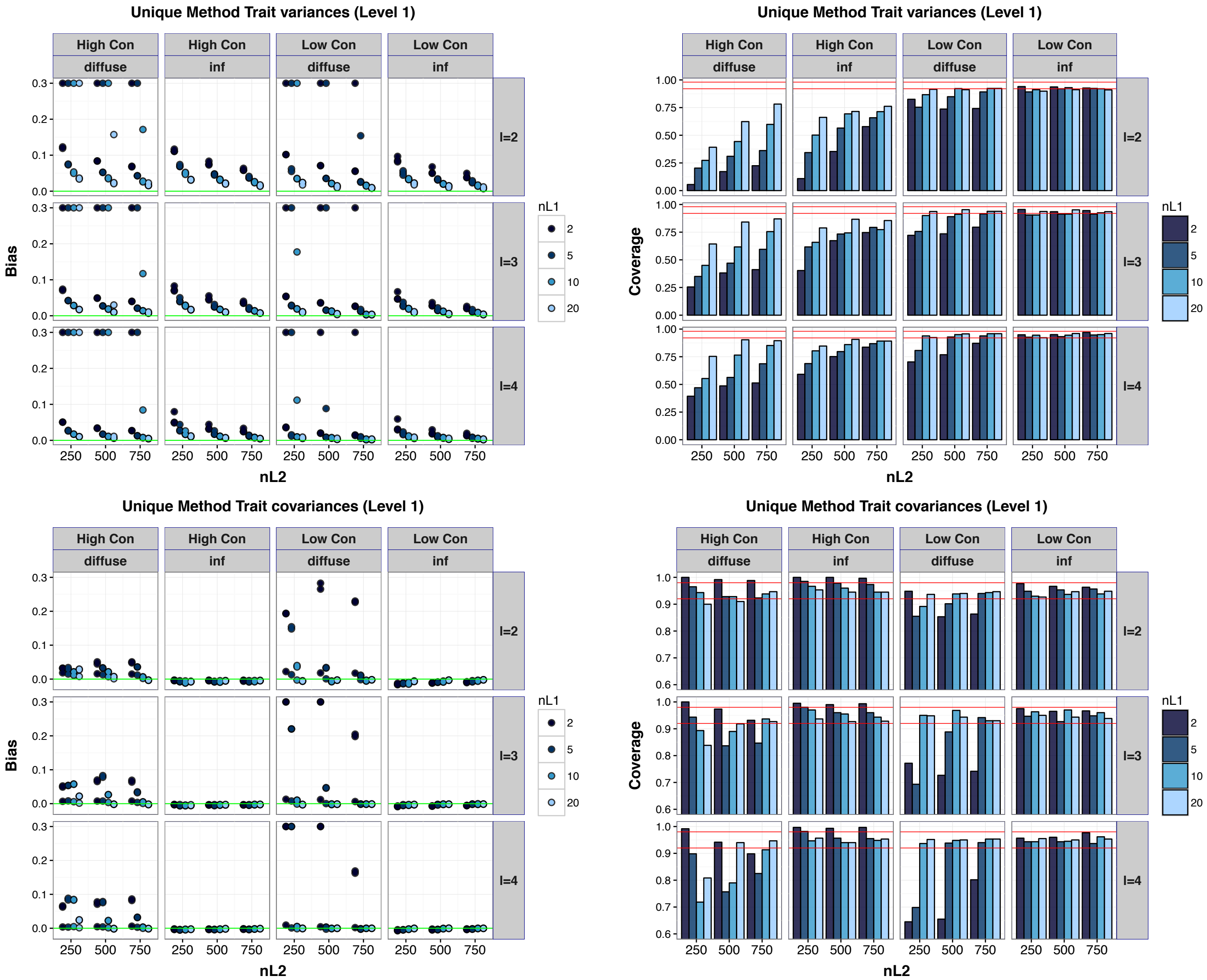

Figure B 13: Bias and 95% coverage for unique method trait factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

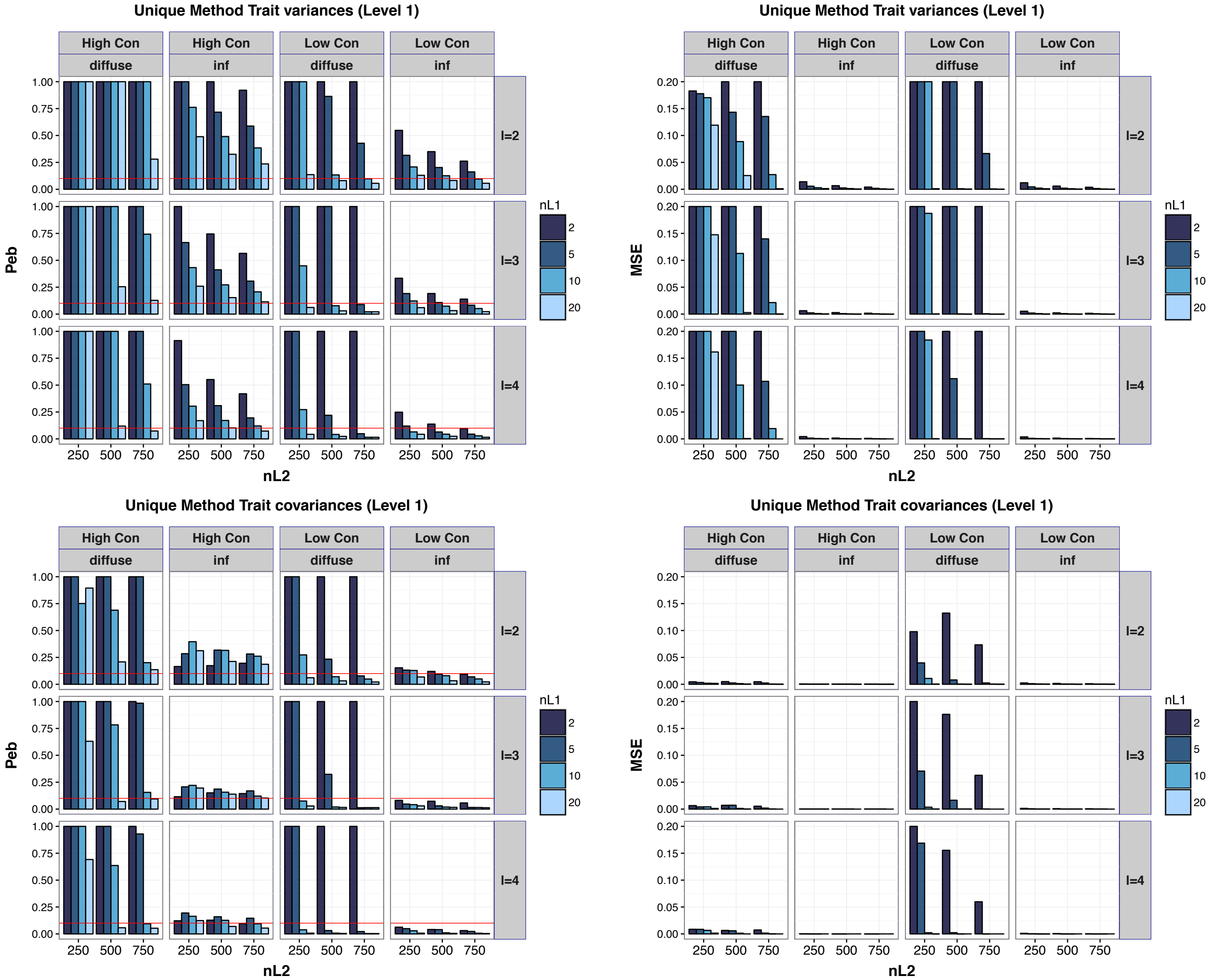

Figure B 14: Parameter estimation bias (peb) and mean squared error (MSE) for unique method trait factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

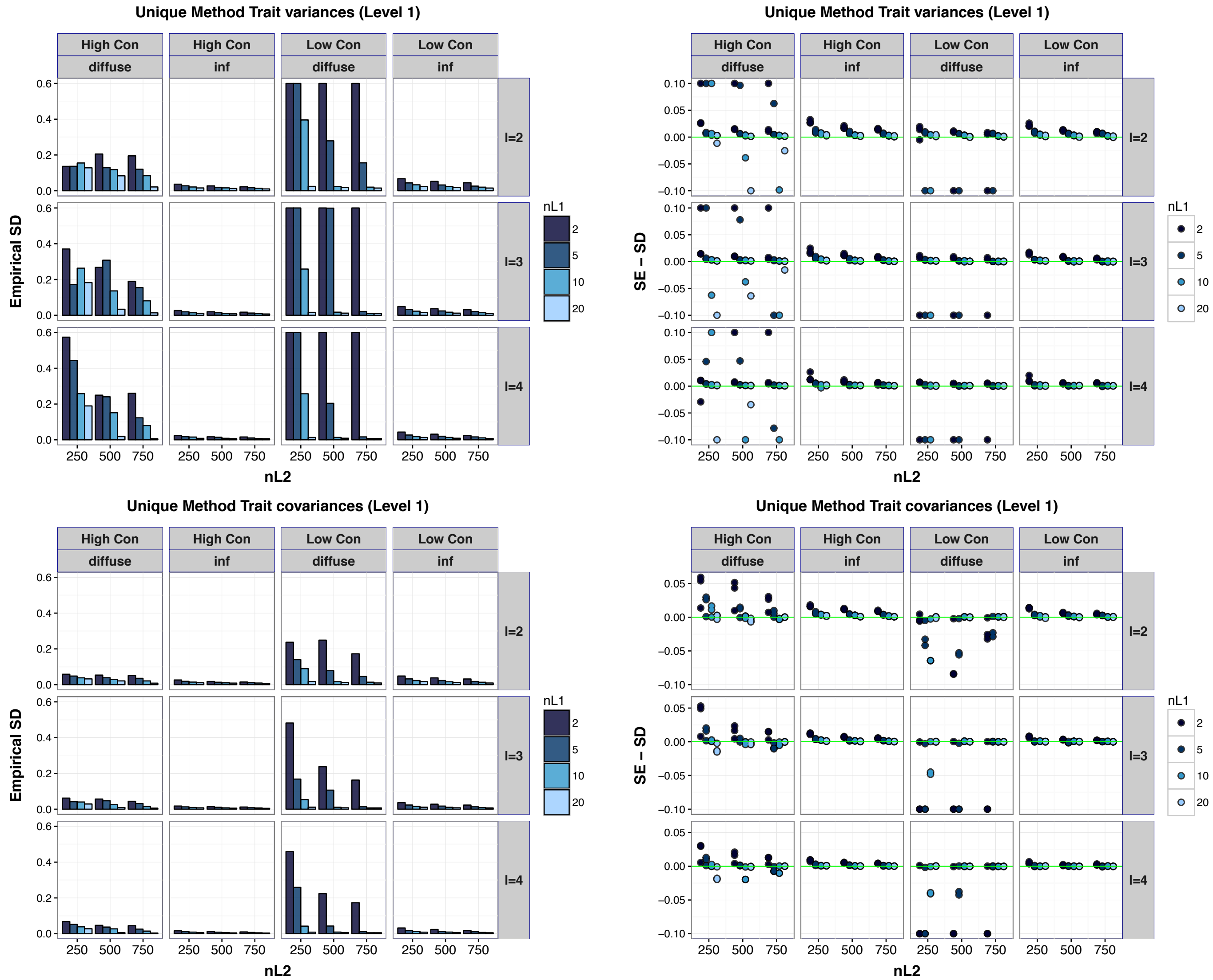

Figure B 15: Empirical SDs and standard error bias (SE - SD) for unique method trait factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

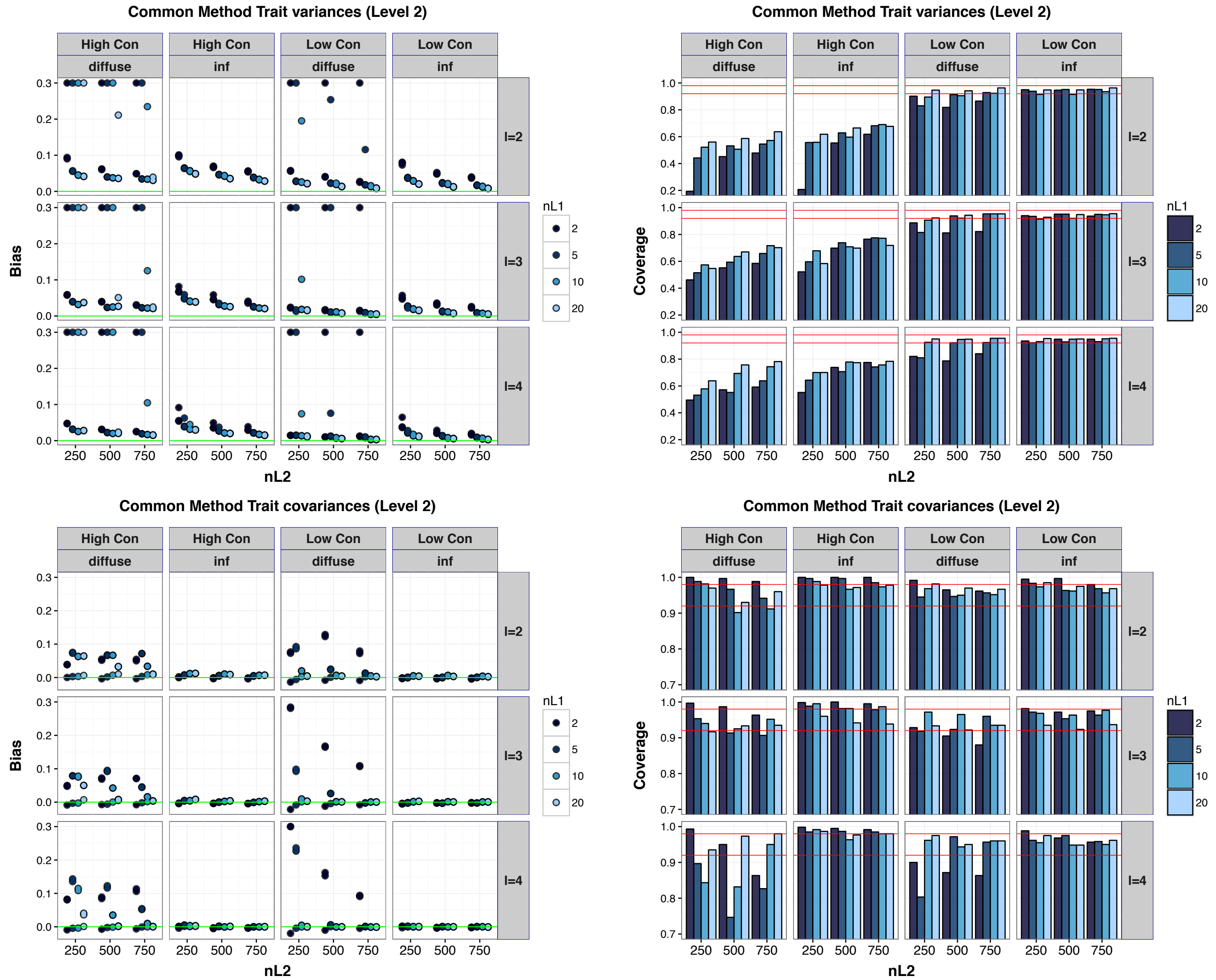

Figure B 16: Bias and 95% coverage for common method trait factors in the LST-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

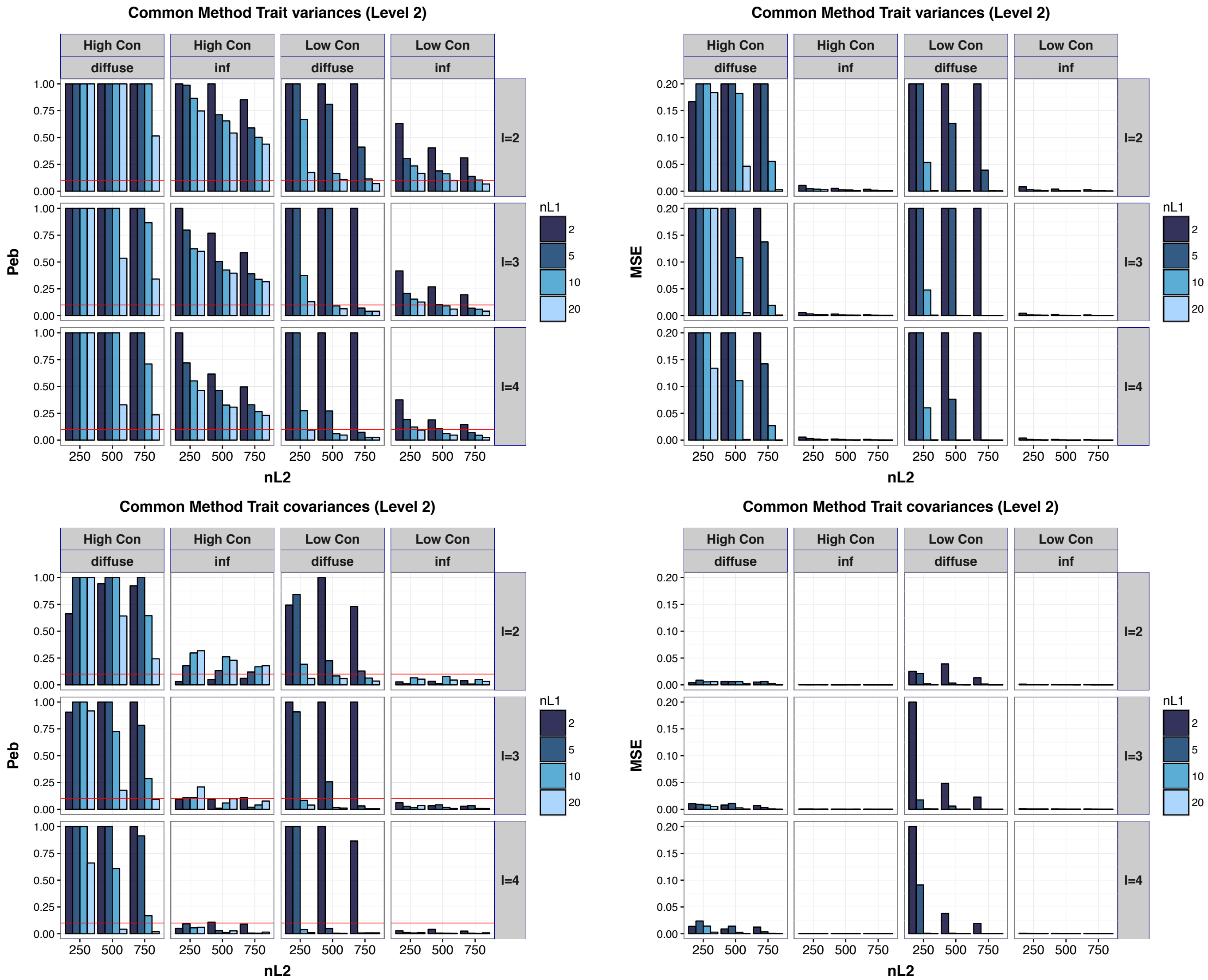

Figure B 17: Parameter estimation bias (peb) and mean squared error (MSE) for common method trait factors in the LST-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

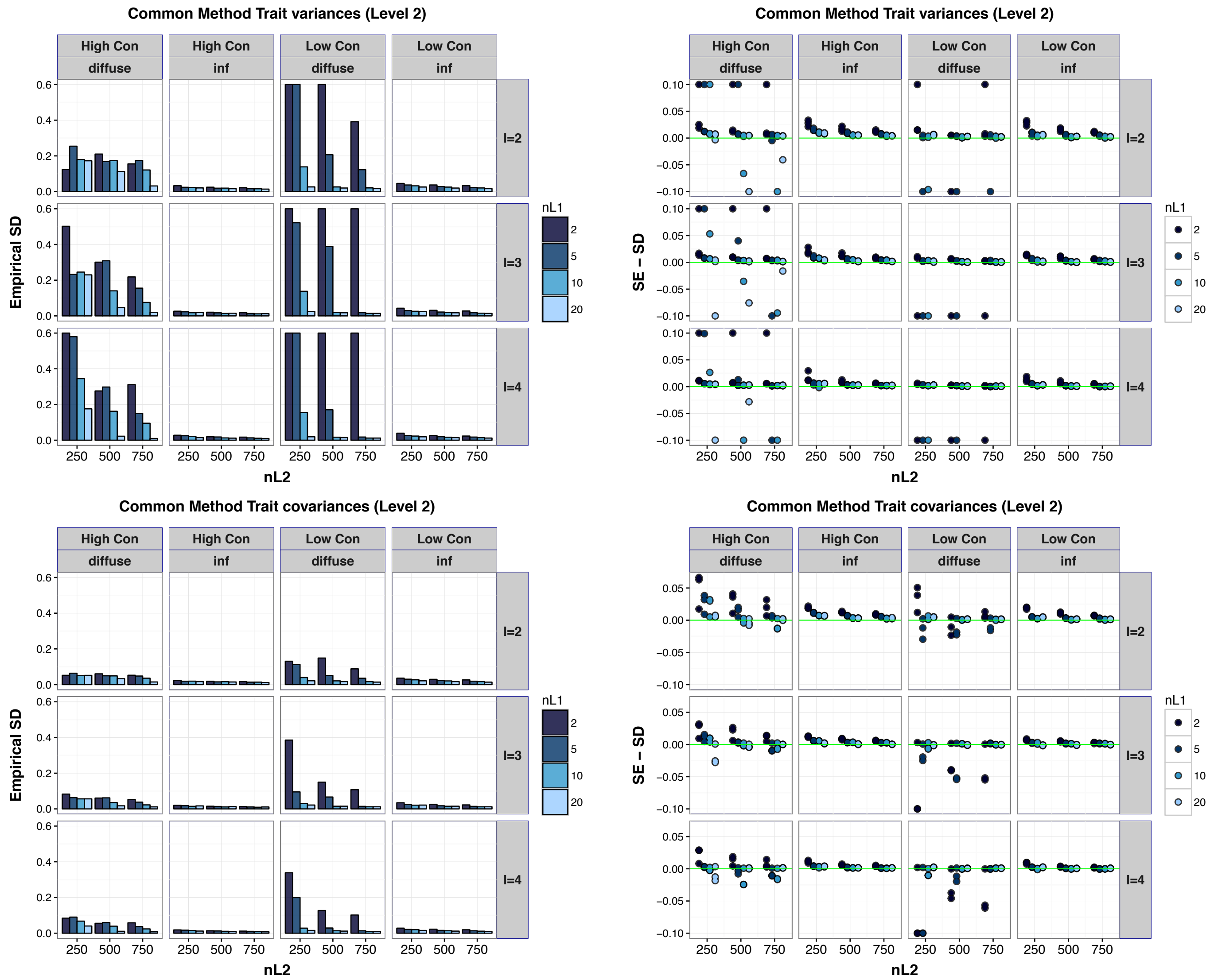

Figure B 18: Empirical SDs and standard error bias (SE - SD) for common method trait factors in the LST-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

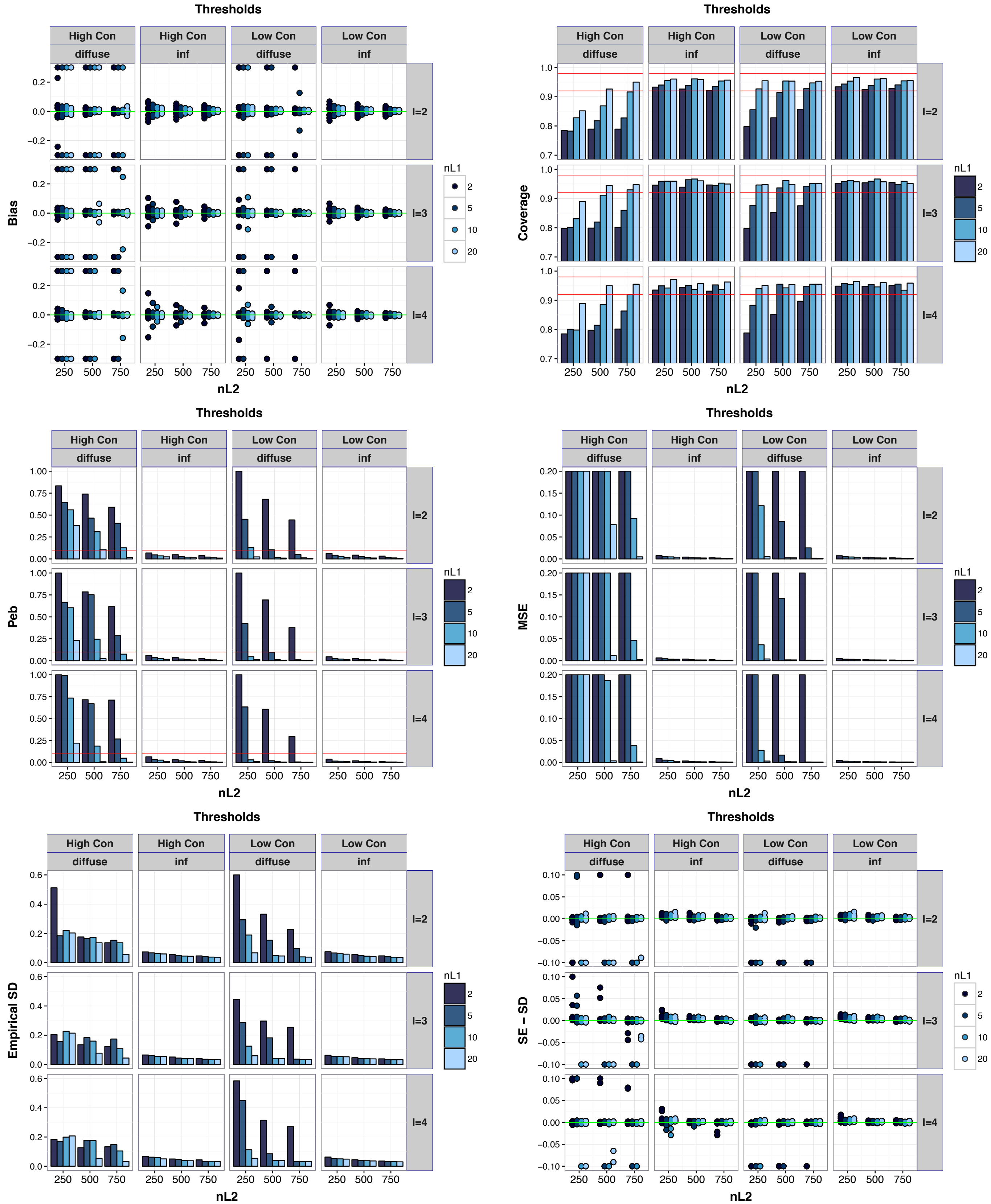

Figure B 19: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LST-Com GRM with one construct. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2, Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

B.3 Simulation results LST-Com GRM. Case of two constructs ($j = 2$)

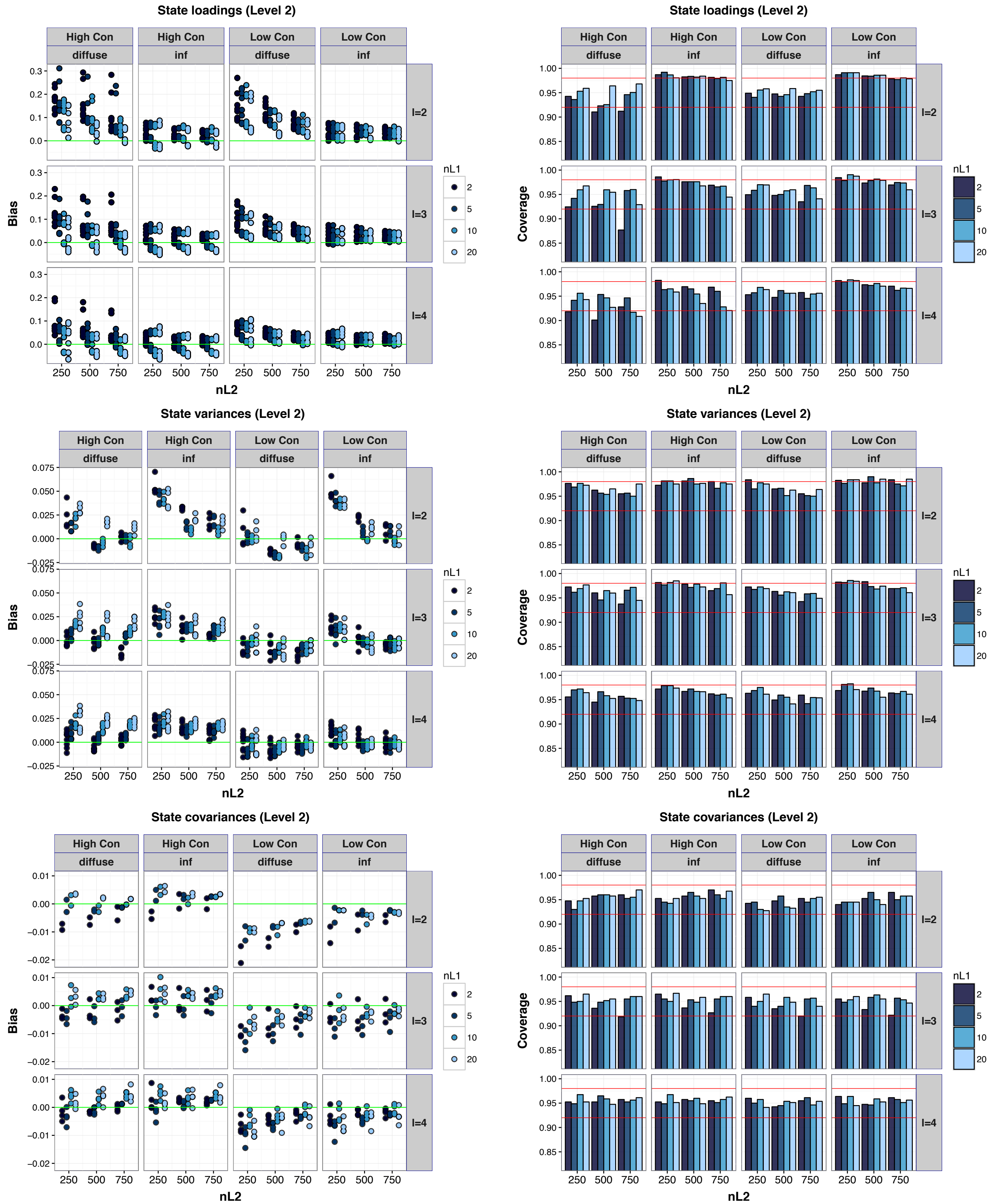

Figure B 20: Bias and 95% coverage for latent state residual factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

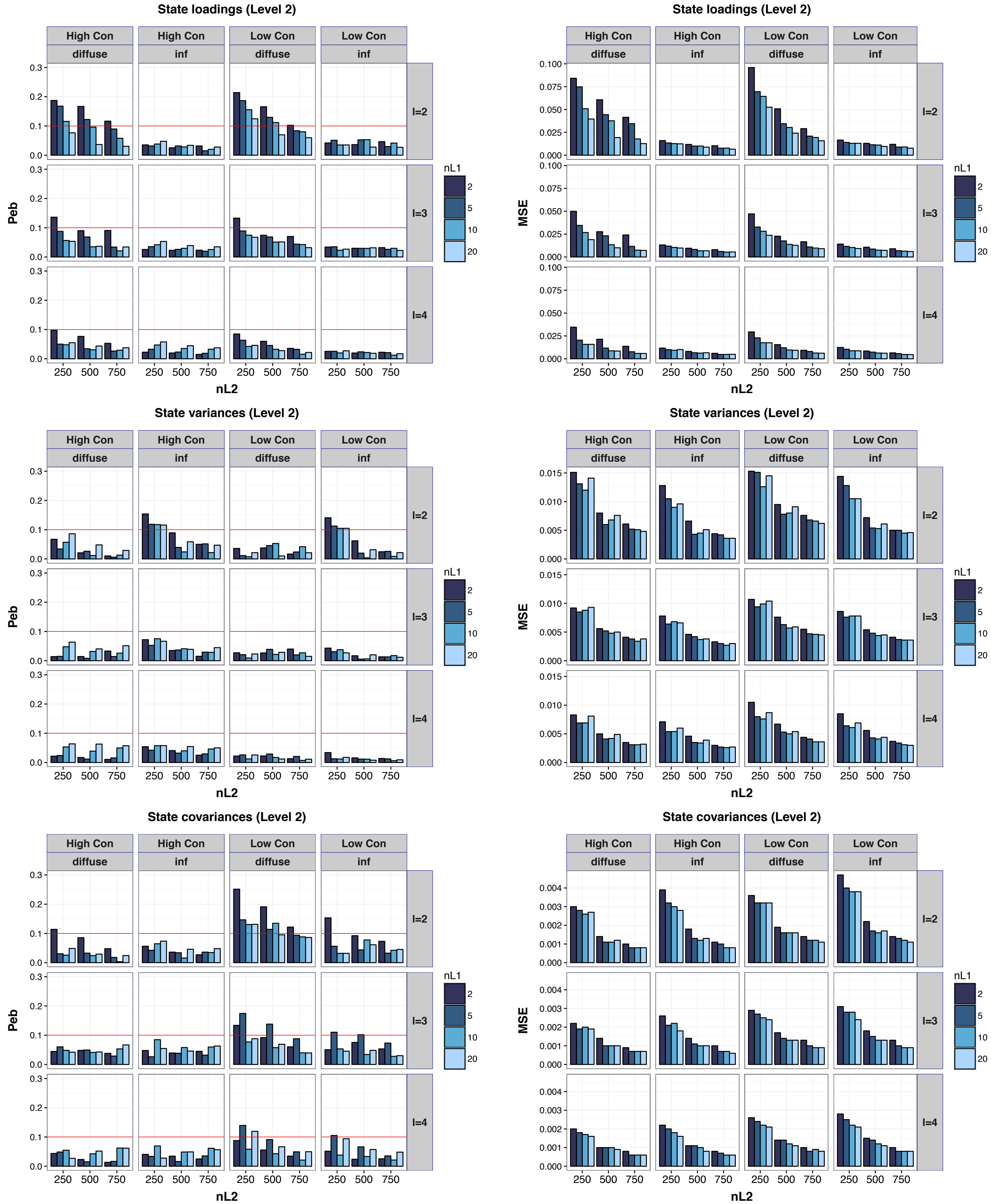

Figure B 21: Parameter estimation bias (peb) and mean squared error (MSE) for latent state residual factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

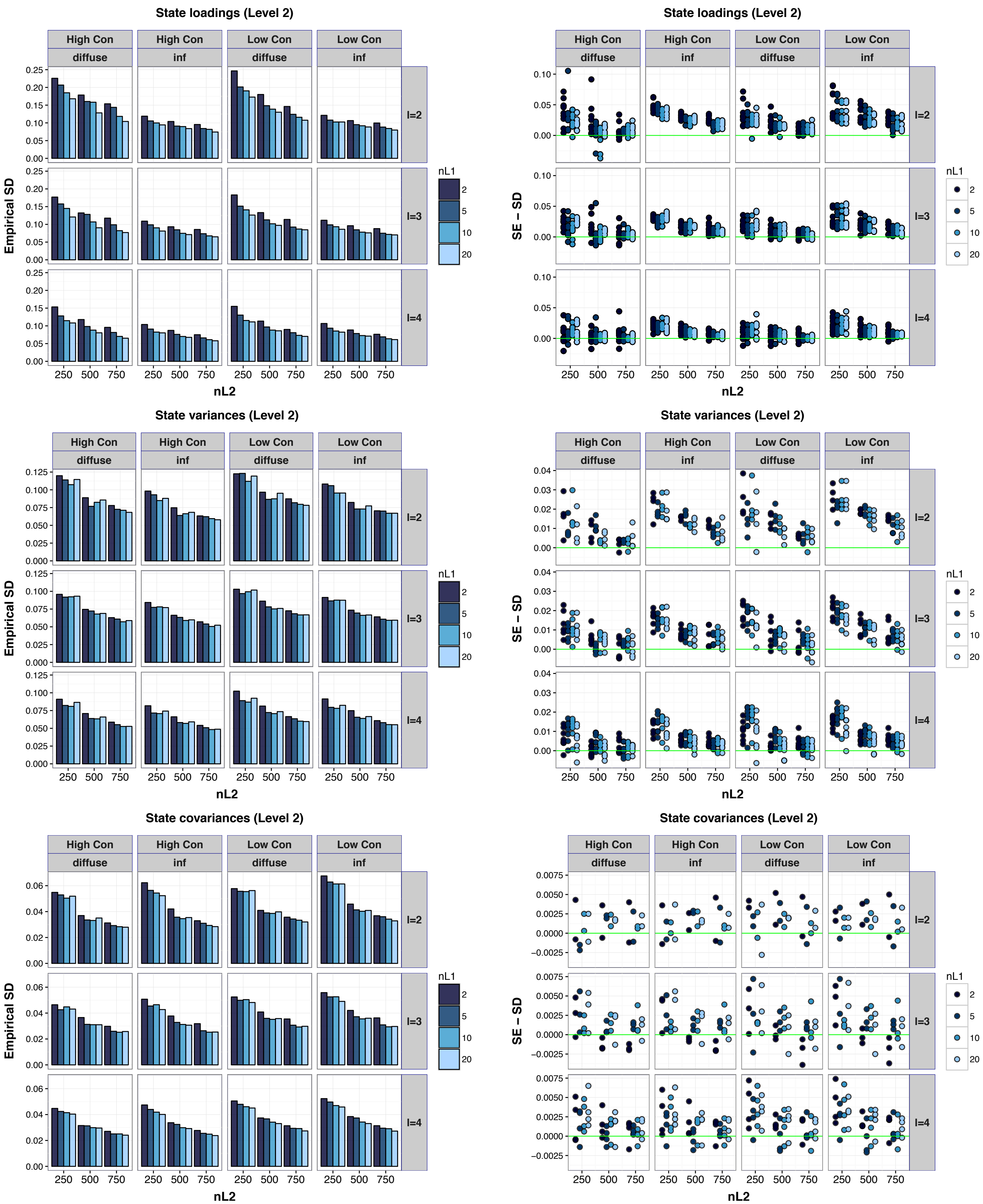

Figure B 22: Empirical SDs and standard error bias (SE - SD) for latent state residual factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

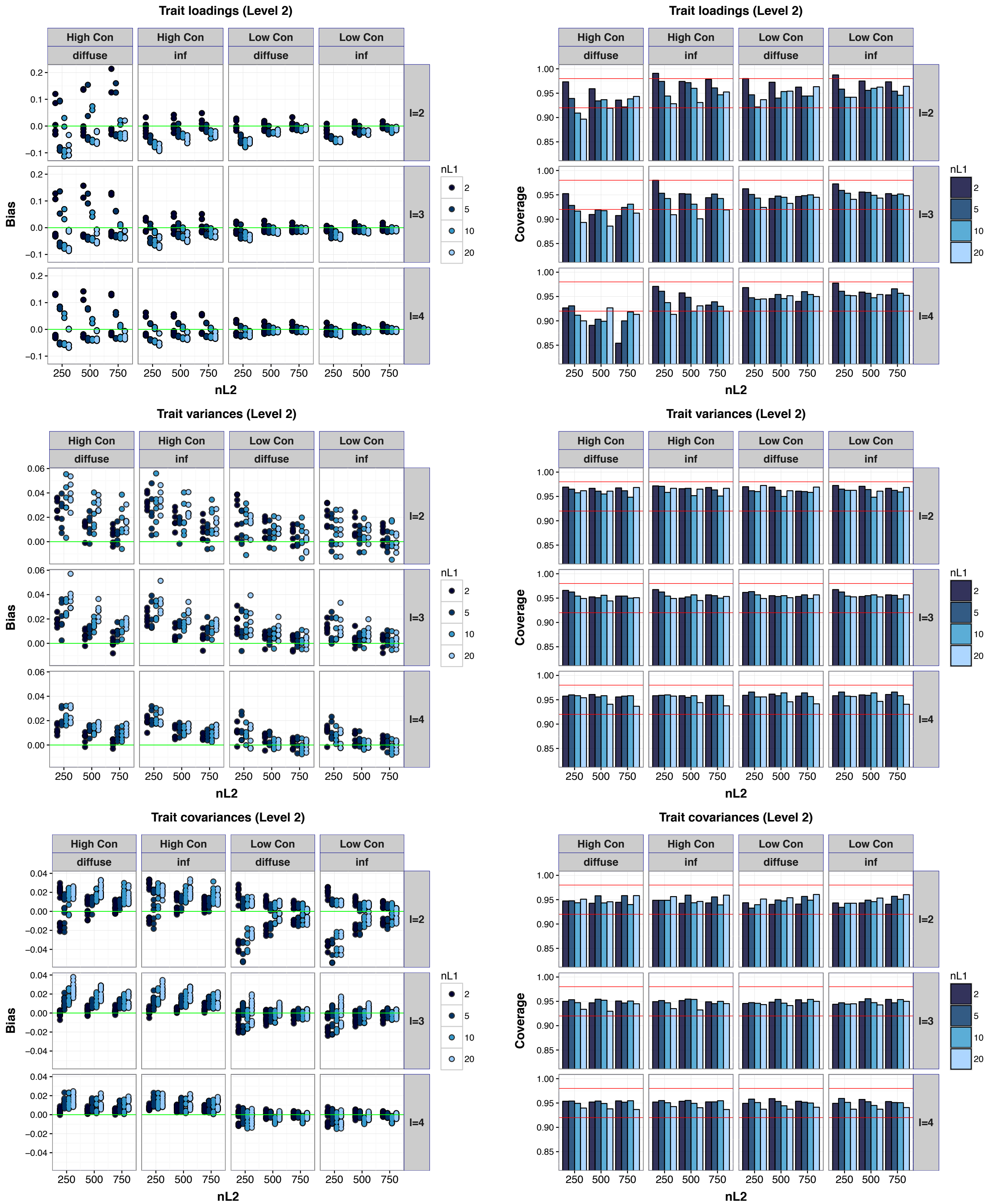

Figure B 23: Bias and 95% coverage for latent trait factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

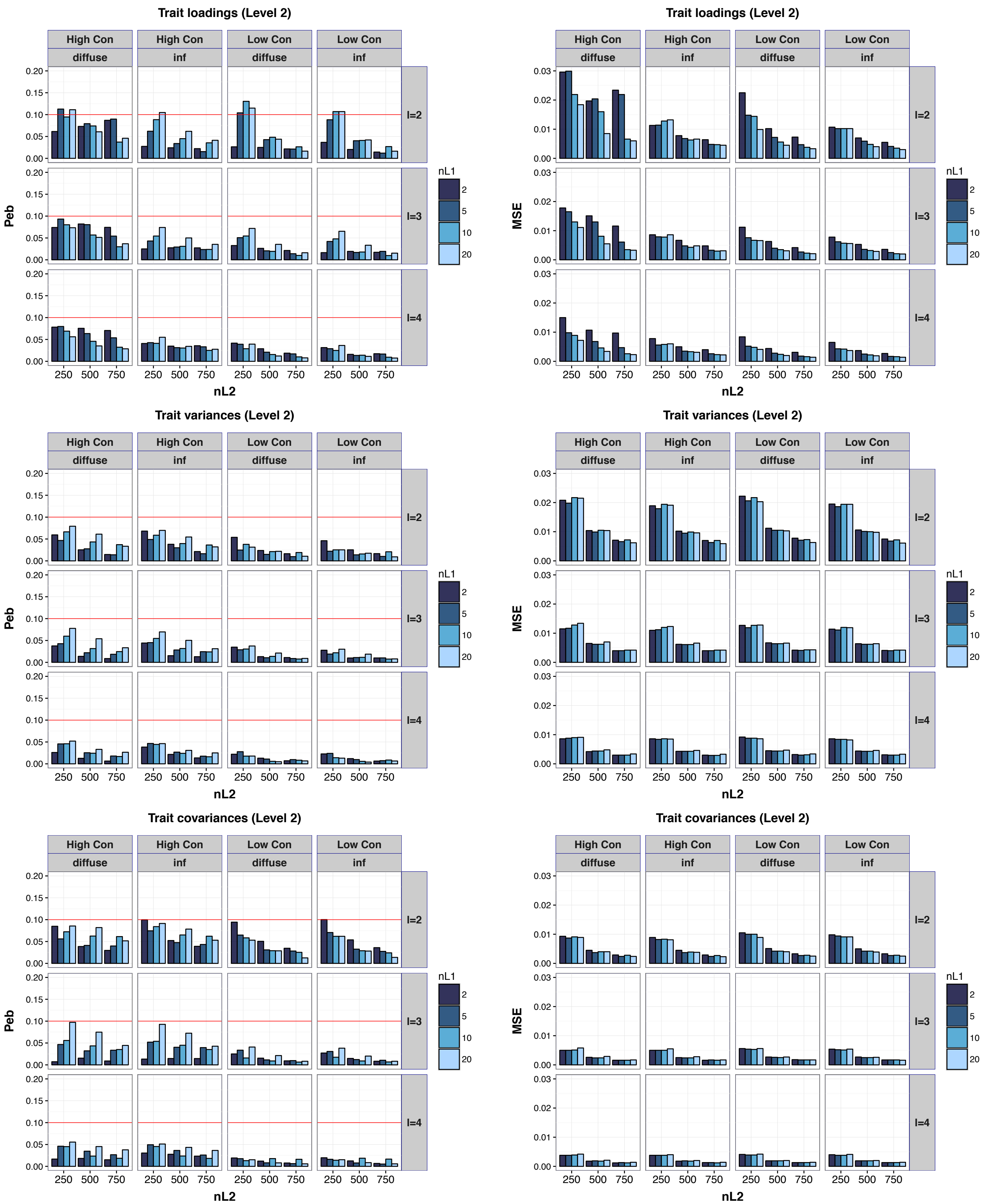

Figure B 24: Parameter estimation bias (peb) and mean squared error (MSE) for latent trait factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

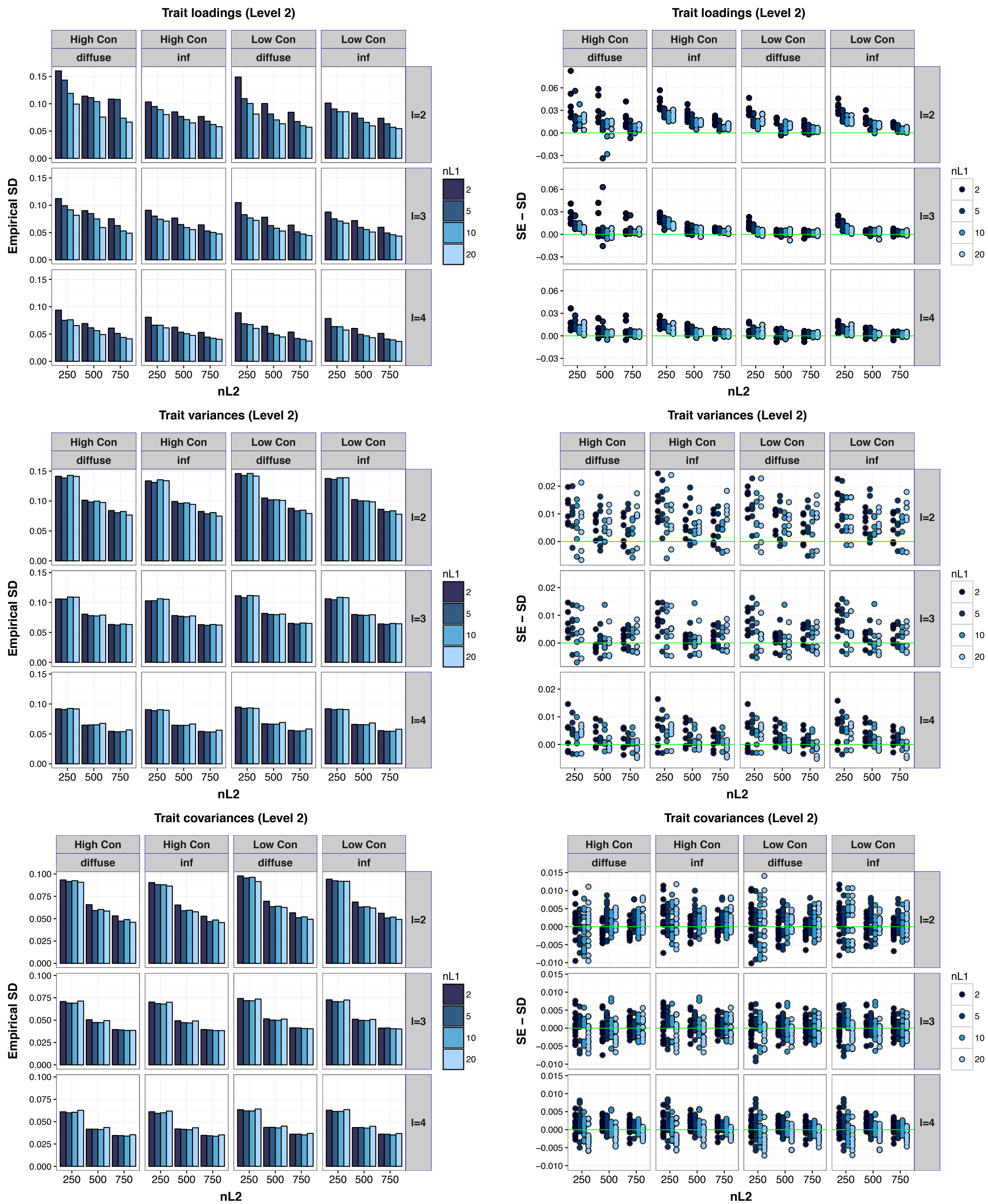

Figure B 25: Empirical SDs and standard error bias (SE - SD) for latent trait factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

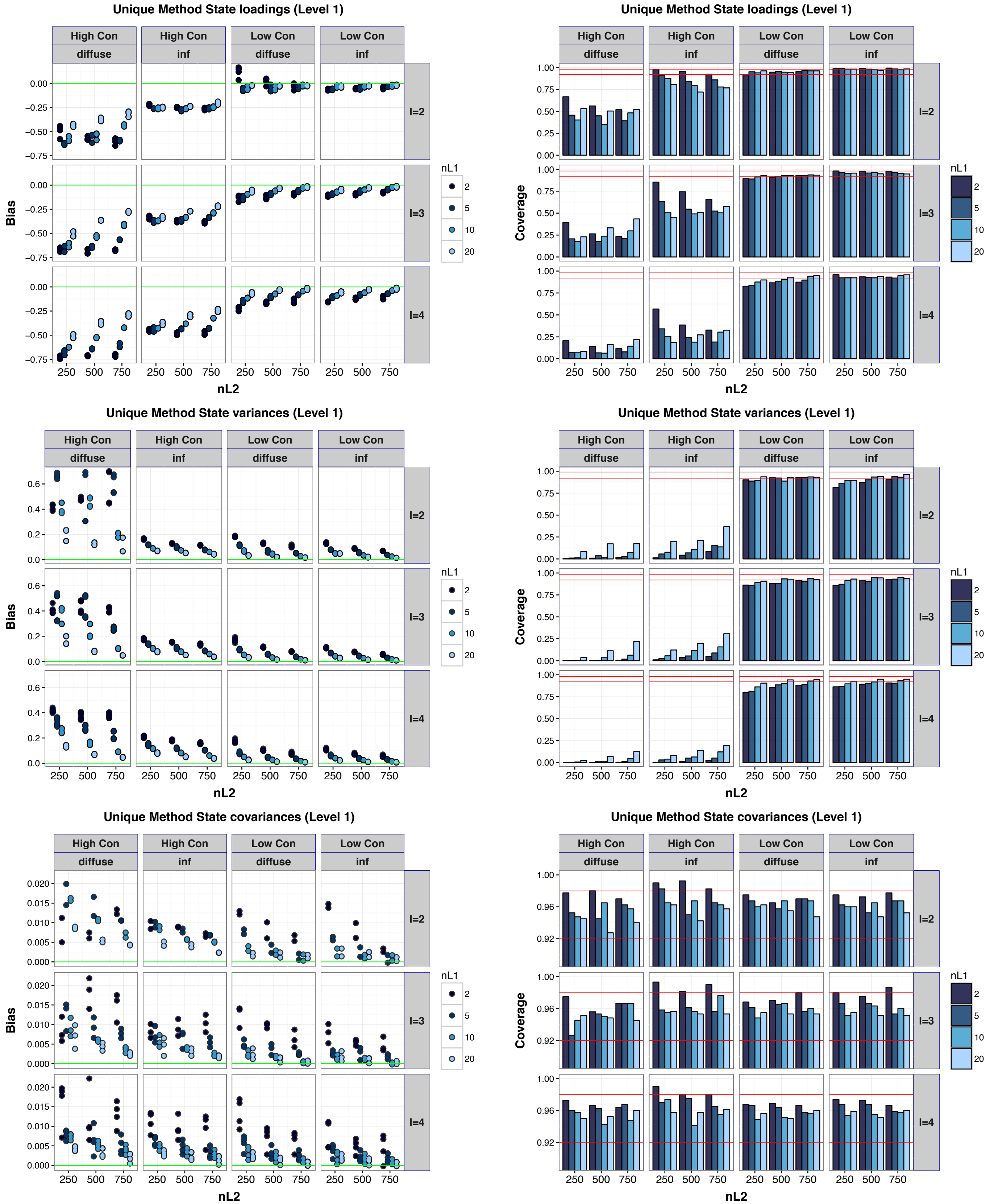

Figure B 26: Bias and 95% coverage for unique method state residual factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

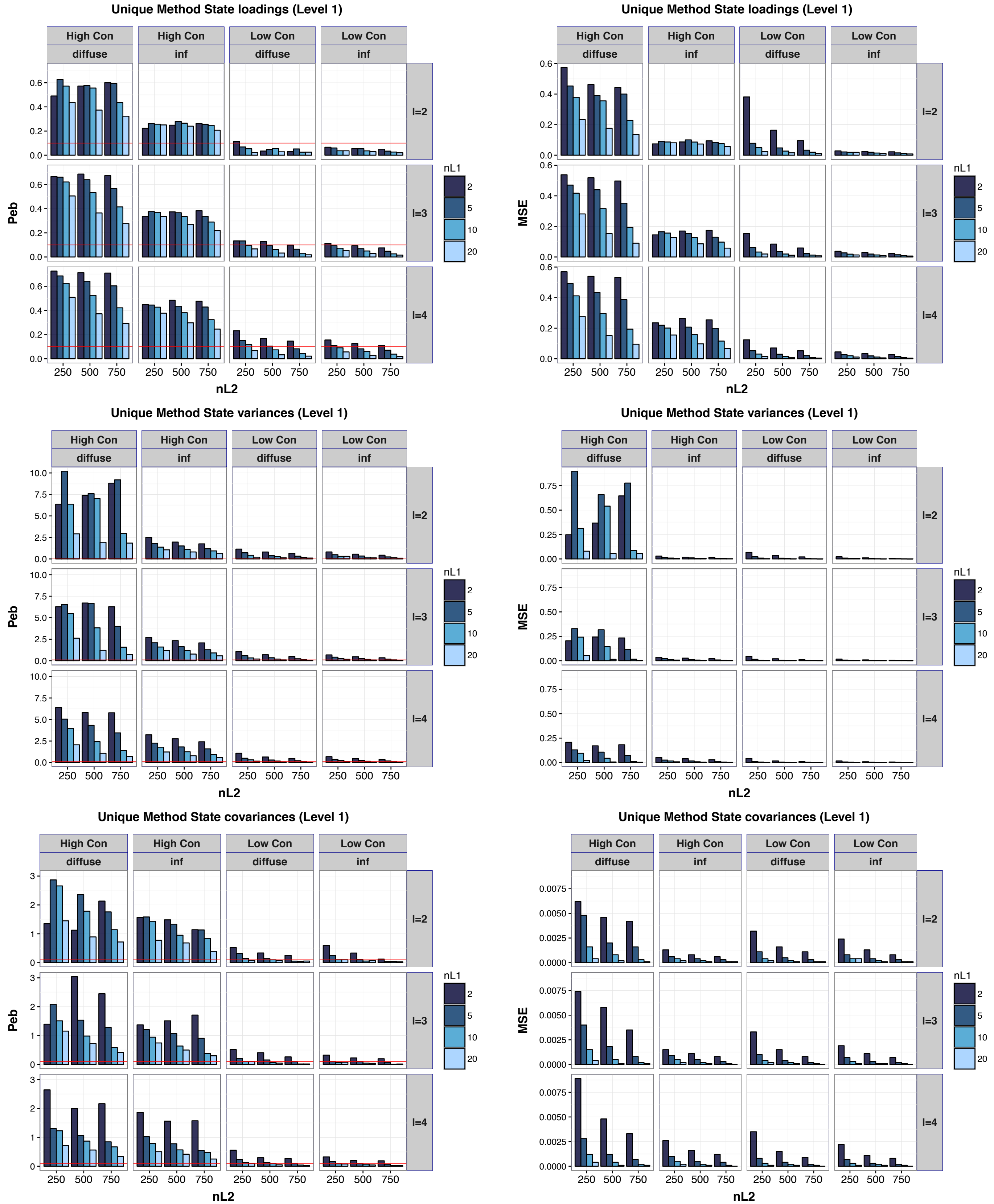

Figure B 27: Parameter estimation bias (peb) and mean squared error (MSE) for unique method state residual factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

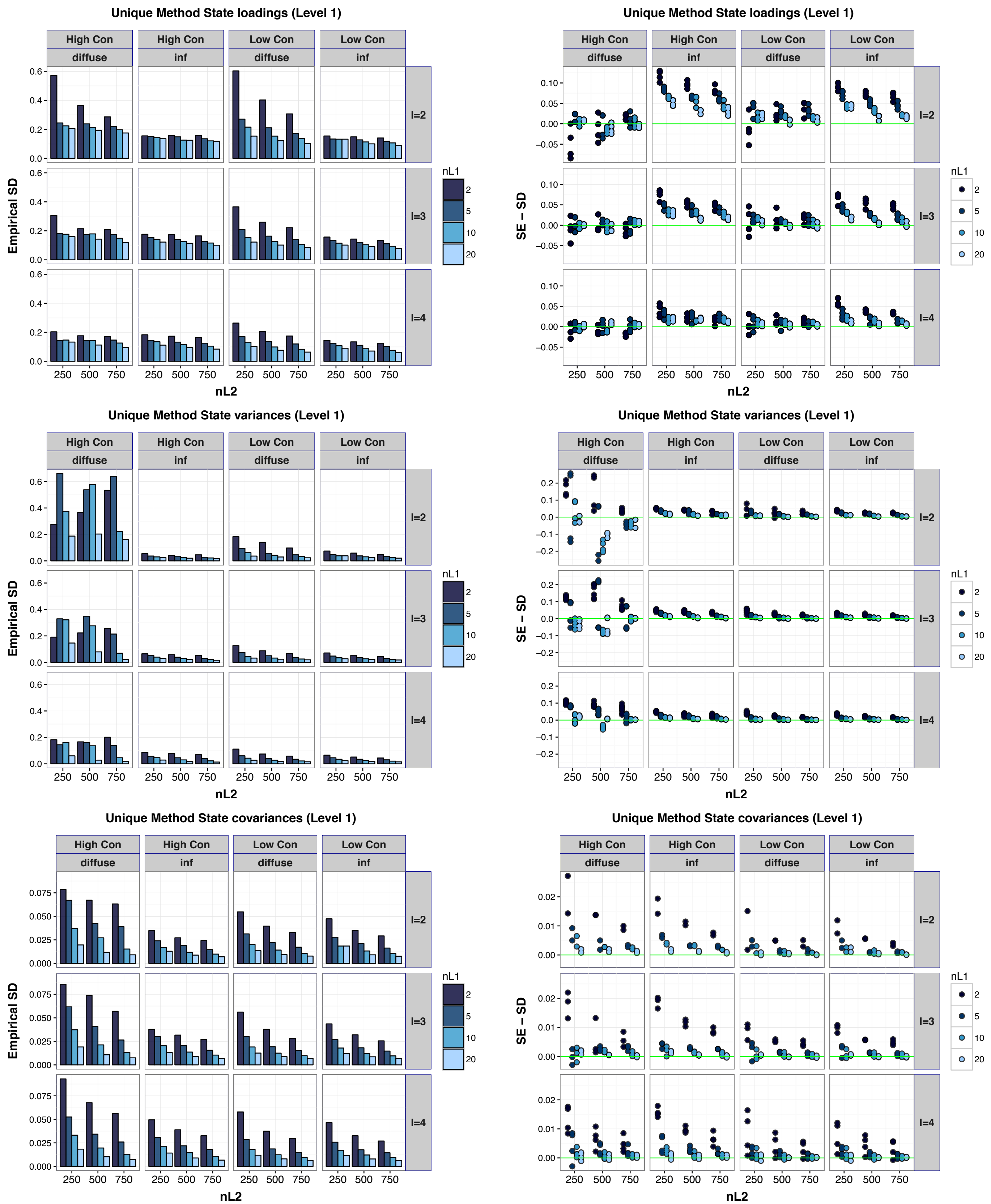

Figure B 28: Empirical SDs and standard error bias (SE - SD) for unique method state residual factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

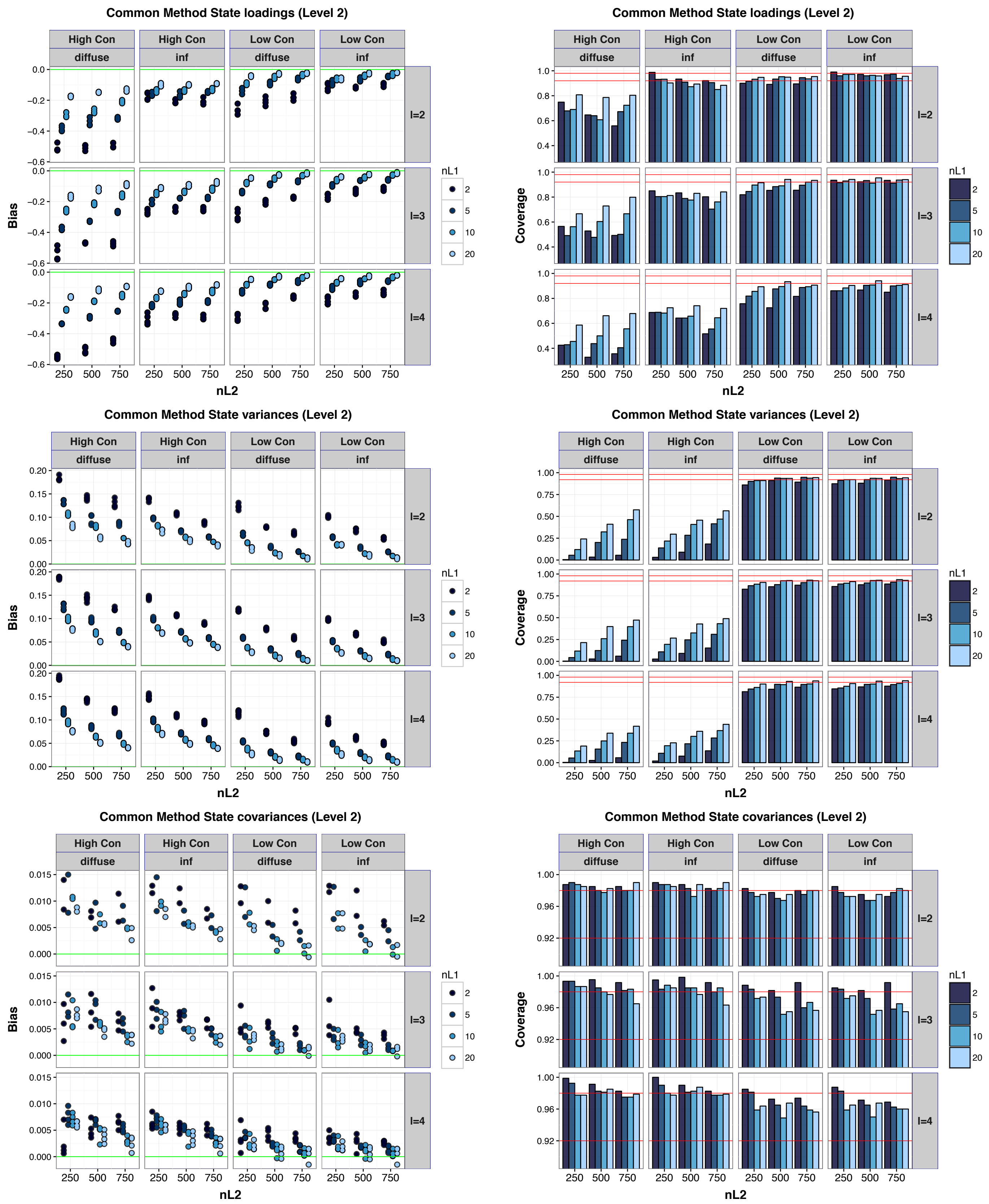

Figure B 29: Bias and 95% coverage for common method state residual factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

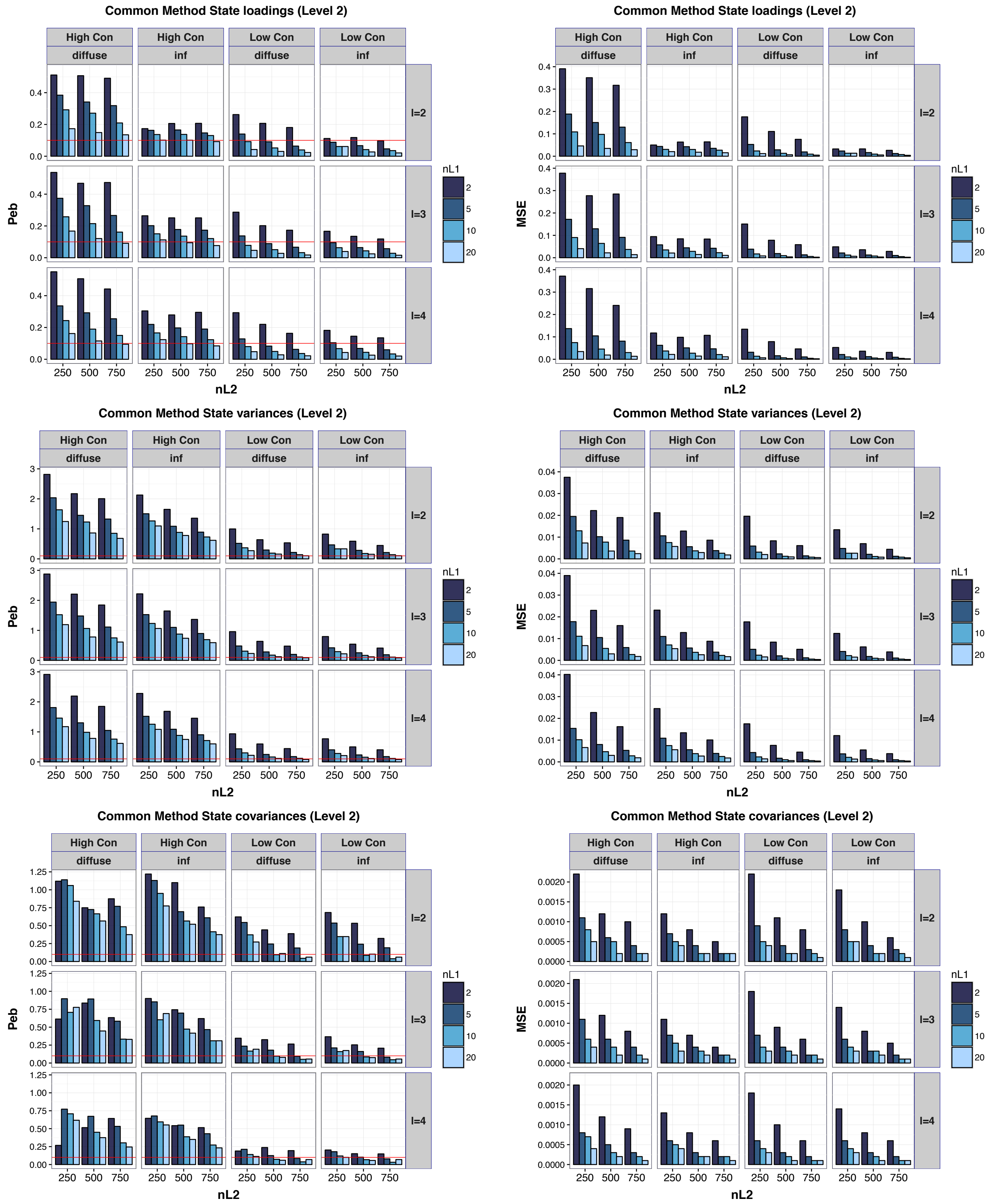

Figure B 30: Parameter estimation bias (peb) and mean squared error (MSE) for common method state residual factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

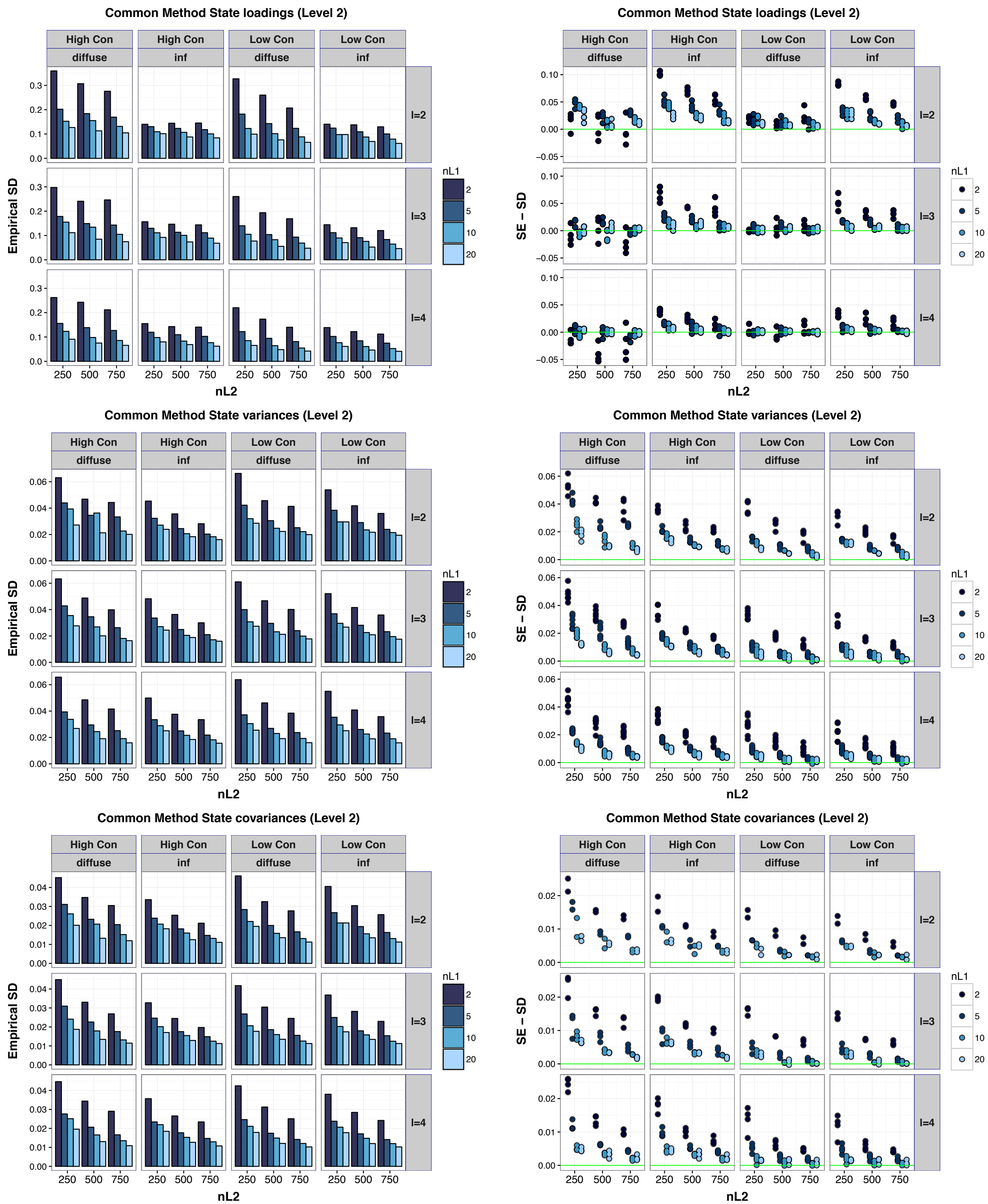

Figure B 31: Empirical SDs and standard error bias (SE - SD) for common method state residual factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

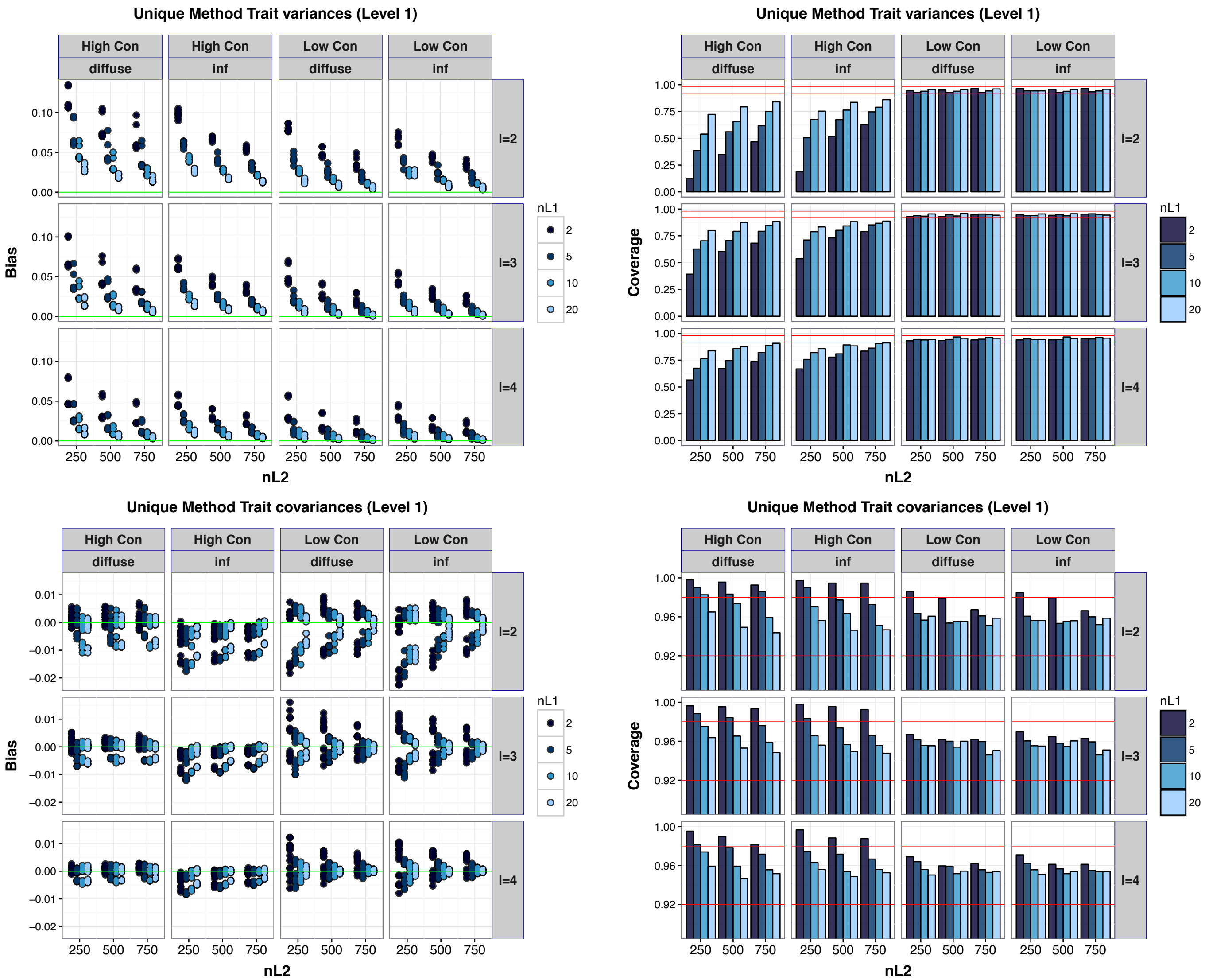

Figure B 32: Bias and 95% coverage for unique method trait factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

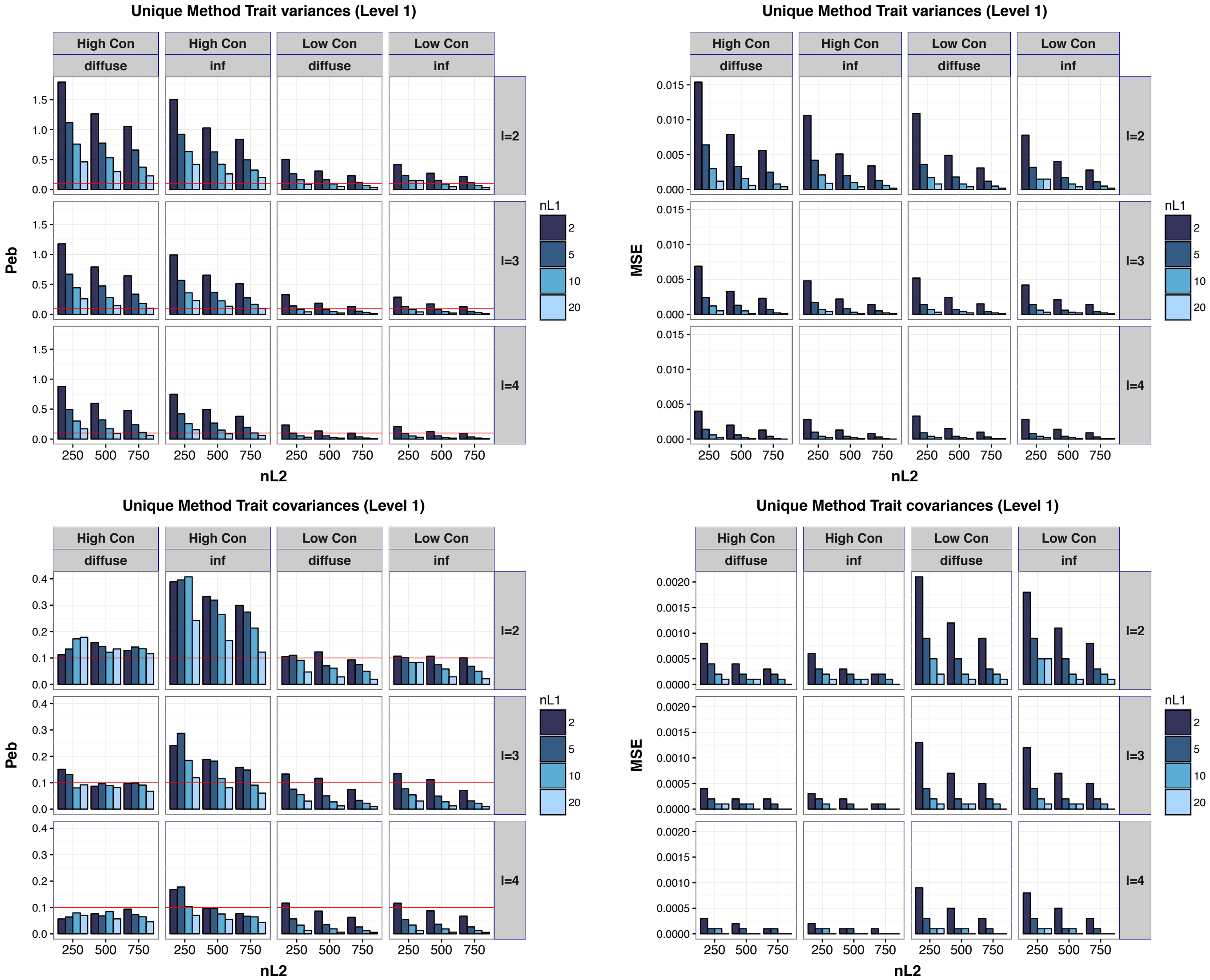

Figure B 33: Parameter estimation bias (peb) and mean squared error (MSE) for unique method trait factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

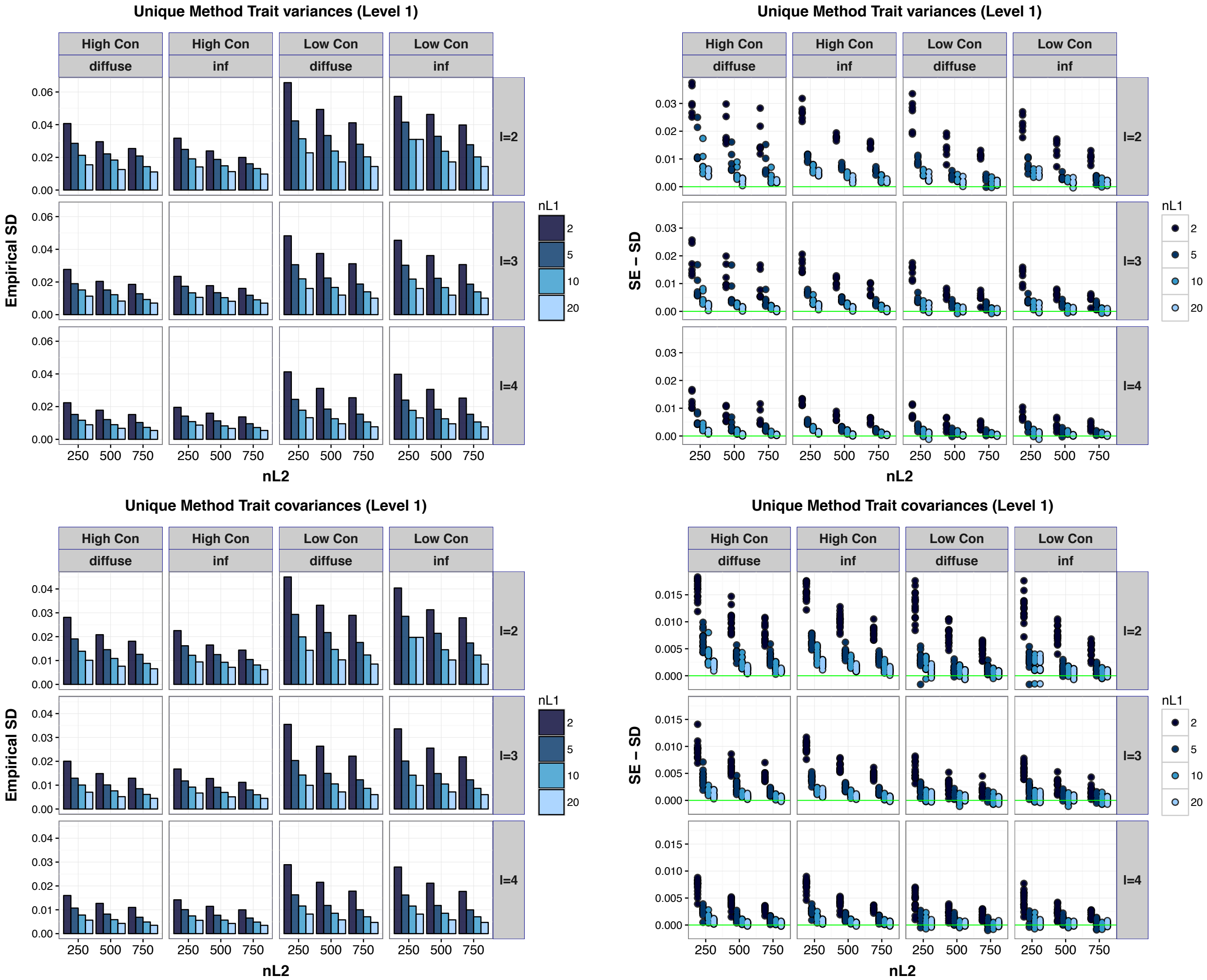

Figure B 34: Empirical SDs and standard error bias (SE - SD) for unique method trait factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

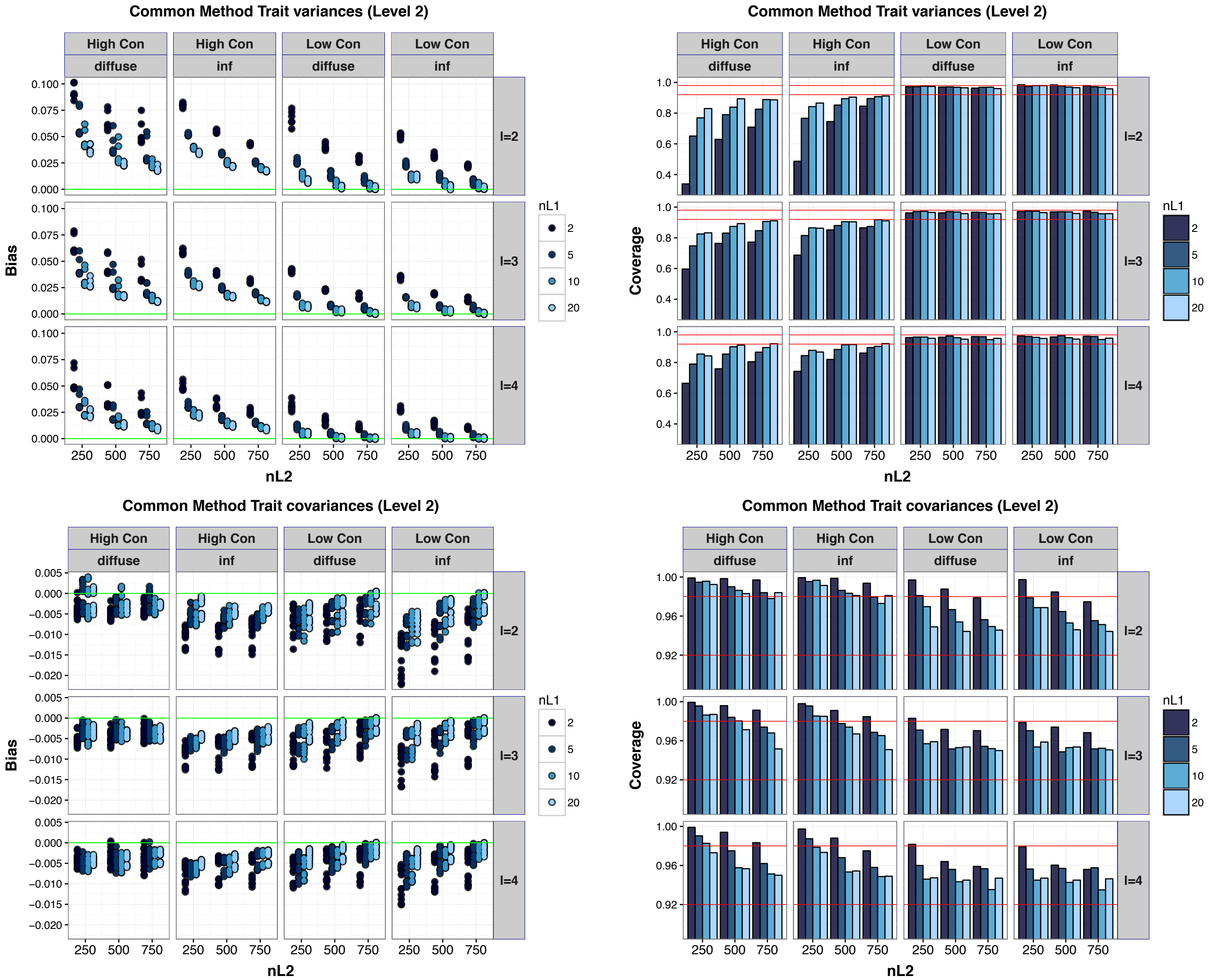

Figure B 35: Bias and 95% coverage for common method trait factors in the LST-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Coverage values were averaged over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

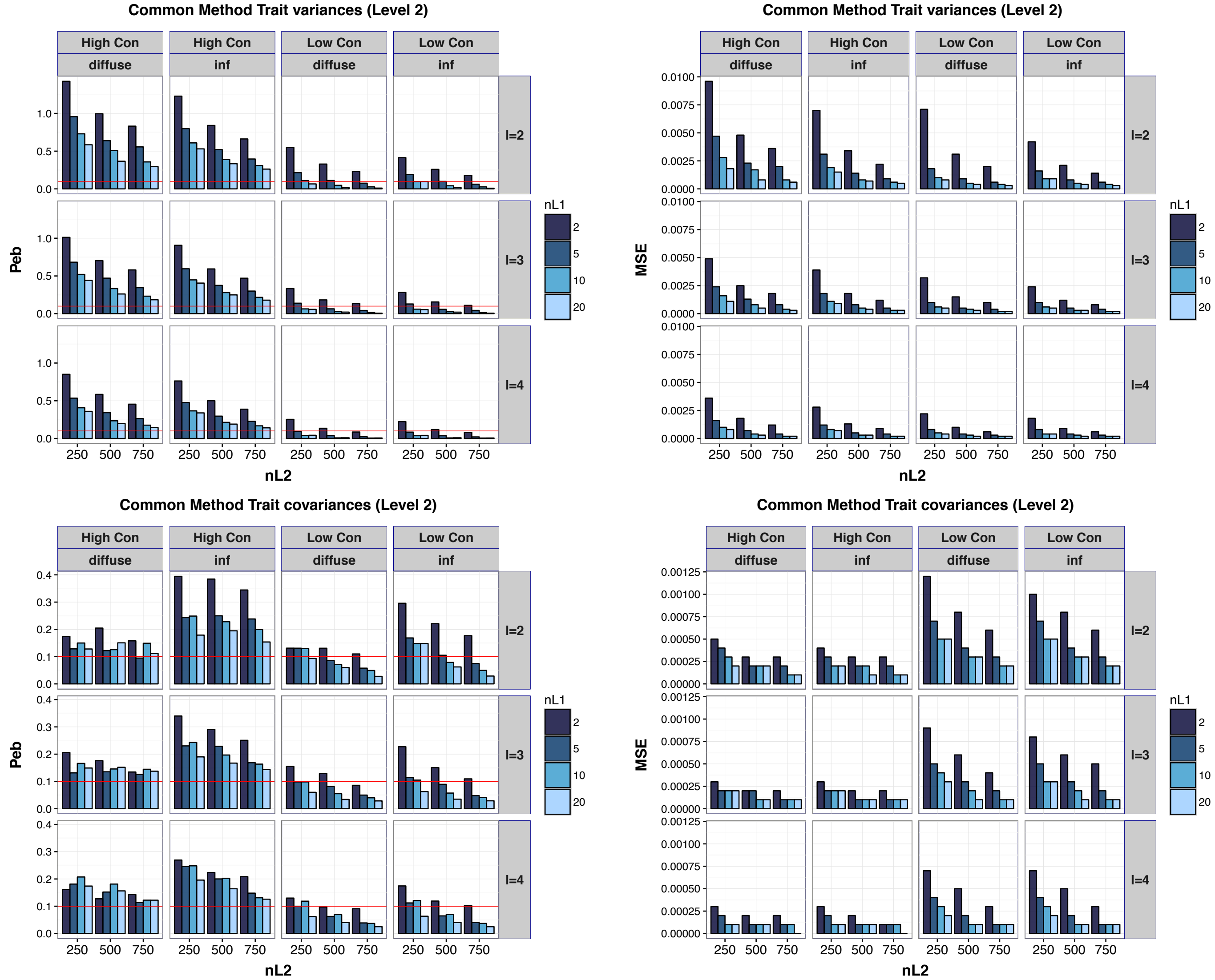

Figure B 36: Parameter estimation bias (peb) and mean squared error (MSE) for common method trait factors in the LST-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

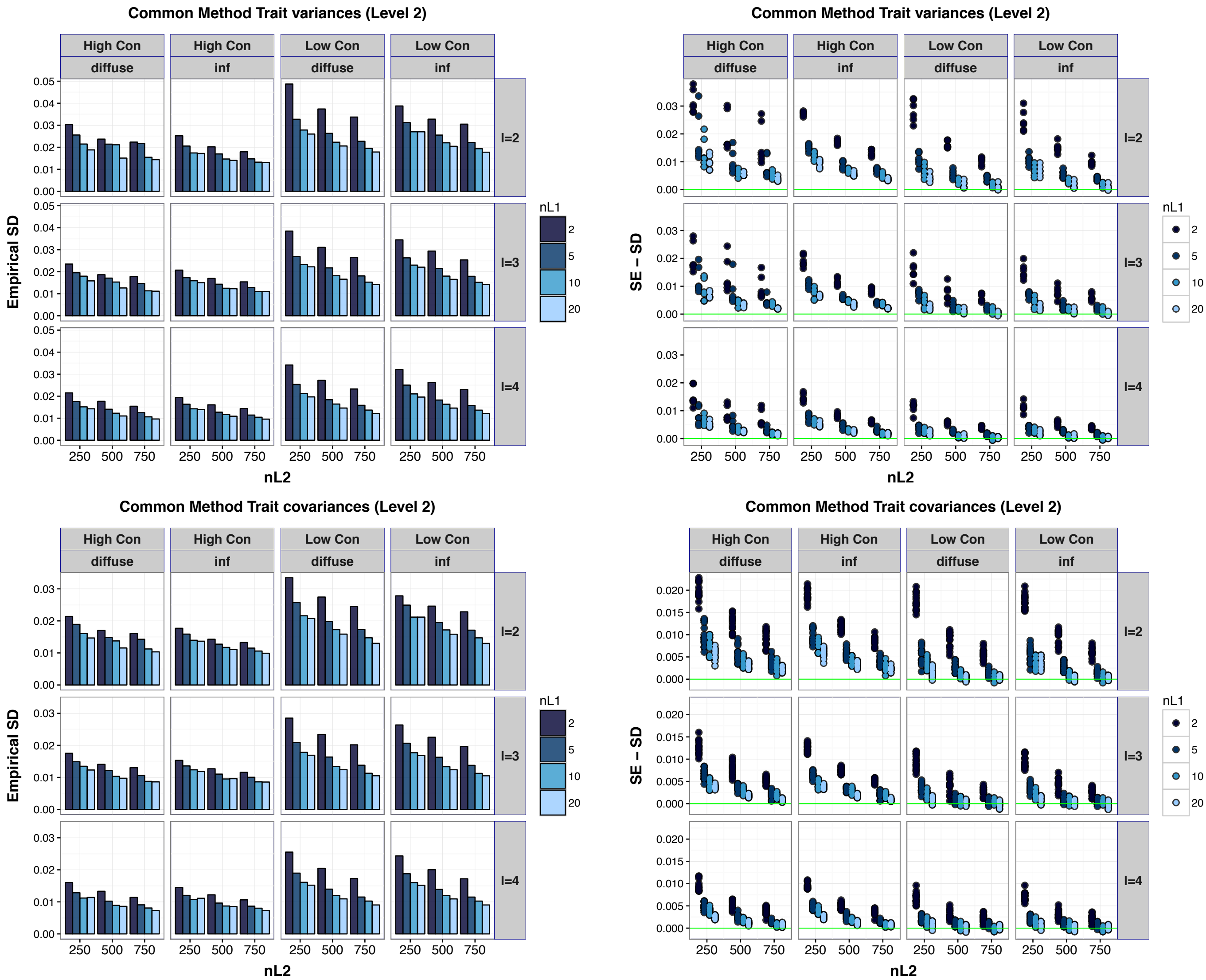

Figure B 37: Empirical SDs and standard error bias (SE - SD) for common method trait factors in the LST-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

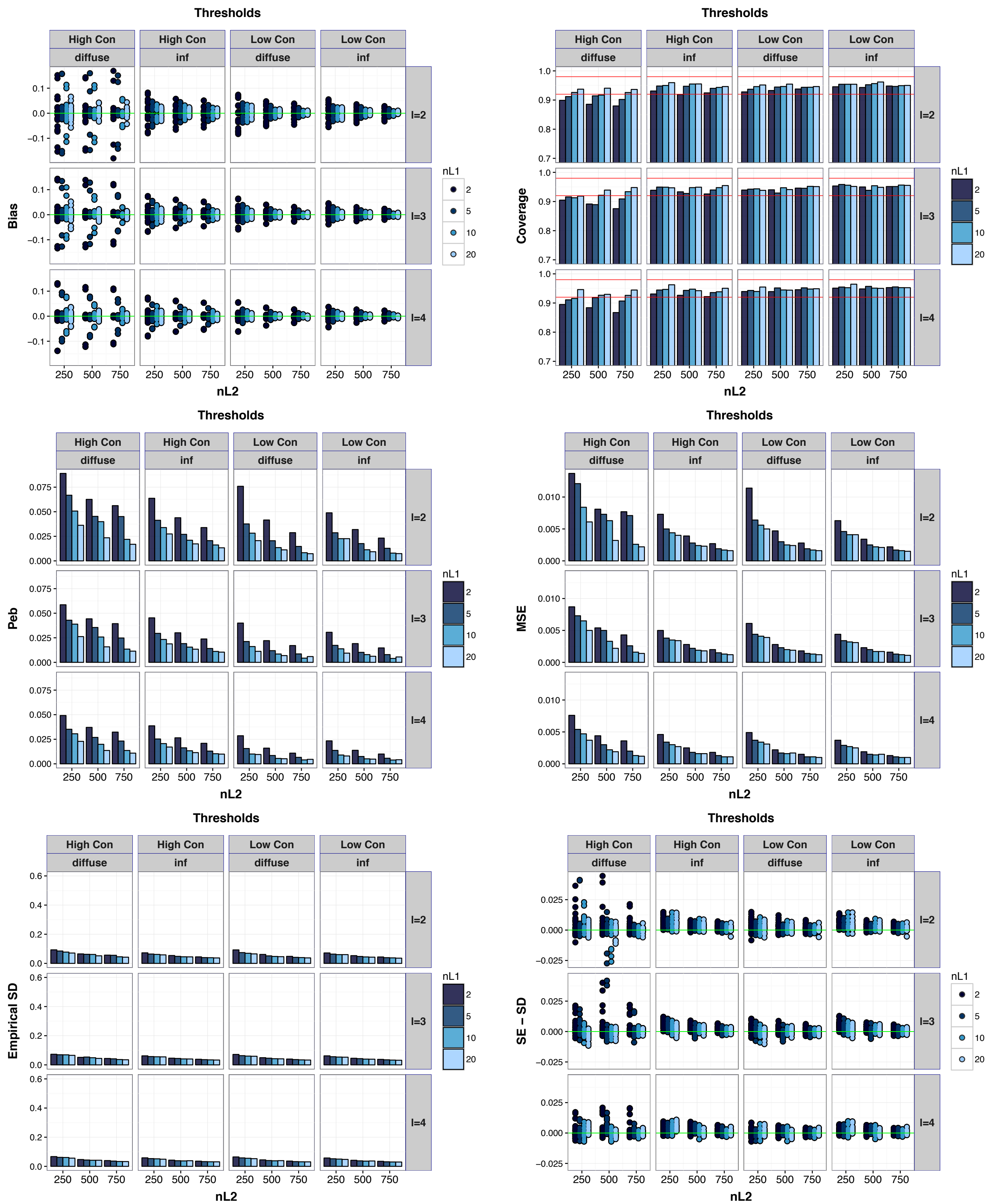

Figure B 38: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LST-Com GRM with two constructs. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2, Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

C Monte Carlo simulation study LGC-Com GRM

C.1 Population parameters in the LGC-Com GRM simulation

Table C 1: Population values in the LGC-Com GRM Monte Carlo simulation study

Population parameters in the LGC-Com GRM simulation				
	Low Con		High Con	
Parameter	$j = 1$	$j = 2$	$j = 1$	$j = 2$
Between-level				
Intercepts				
Loadings	0.5	0.5	0.786	0.786
Variances	0.452	0.452	0.452	0.452
Covariances	0.361	0.181	0.361	0.181
Slopes				
Loadings	0.5	0.5	0.786	0.786
Variances	0.038	0.038	0.038	0.038
Covariances	0.031	0.015	0.031	0.015
Means	0.4	0.4	0.4	0.4
Common Method Intercepts				
Variances	0.123	0.123	0.065	0.065
Covariances	0.074	0.037	0.039	0.020

Note. Values of the population parameters that differ from those in the LST-Com GRM simulation. Note that population values for within-level parameters (i.e., unique method trait and state residual loadings, variances, and covariances) as well as those of state residual and common method state residual variables are identical to those in the LST-Com GRM simulation study (see Table B 1) and are not reported here. Population values did not vary between the parameters of one parameter class. Note that covariance parameters reported for the multi-construct conditions ($j = 2$) correspond to the covariances of the respective factors between constructs. Covariance within one construct are identical in the mono-construct and multi-construct conditions and reported under $j = 1$. High Con; high consistency condition; j: number of constructs; Low Con: Low consistency condition.

C.2 Simulation results LGC-Com GRM. Case of one construct ($j = 1$).

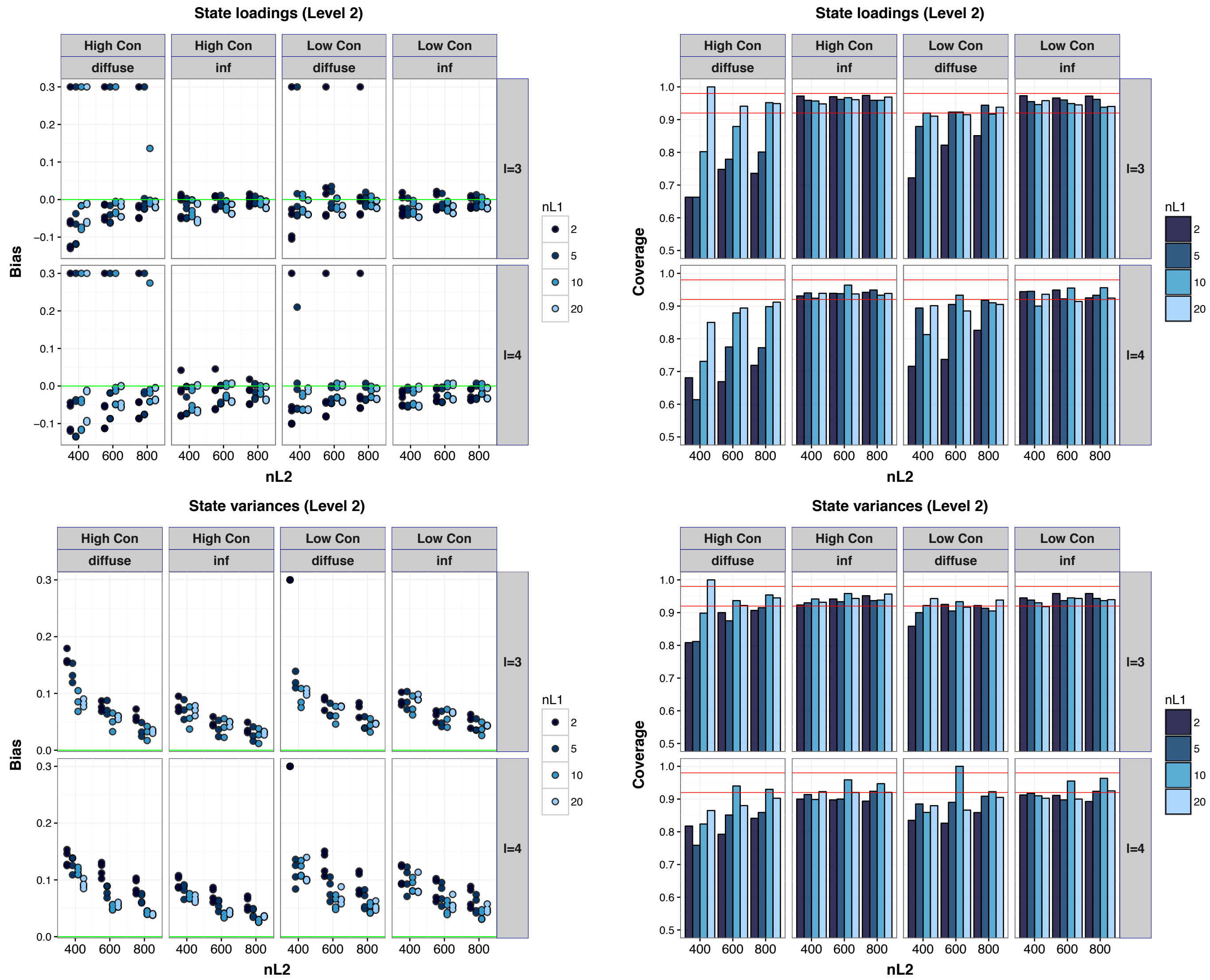

Figure C 1: Bias and 95% coverage for latent state residual factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

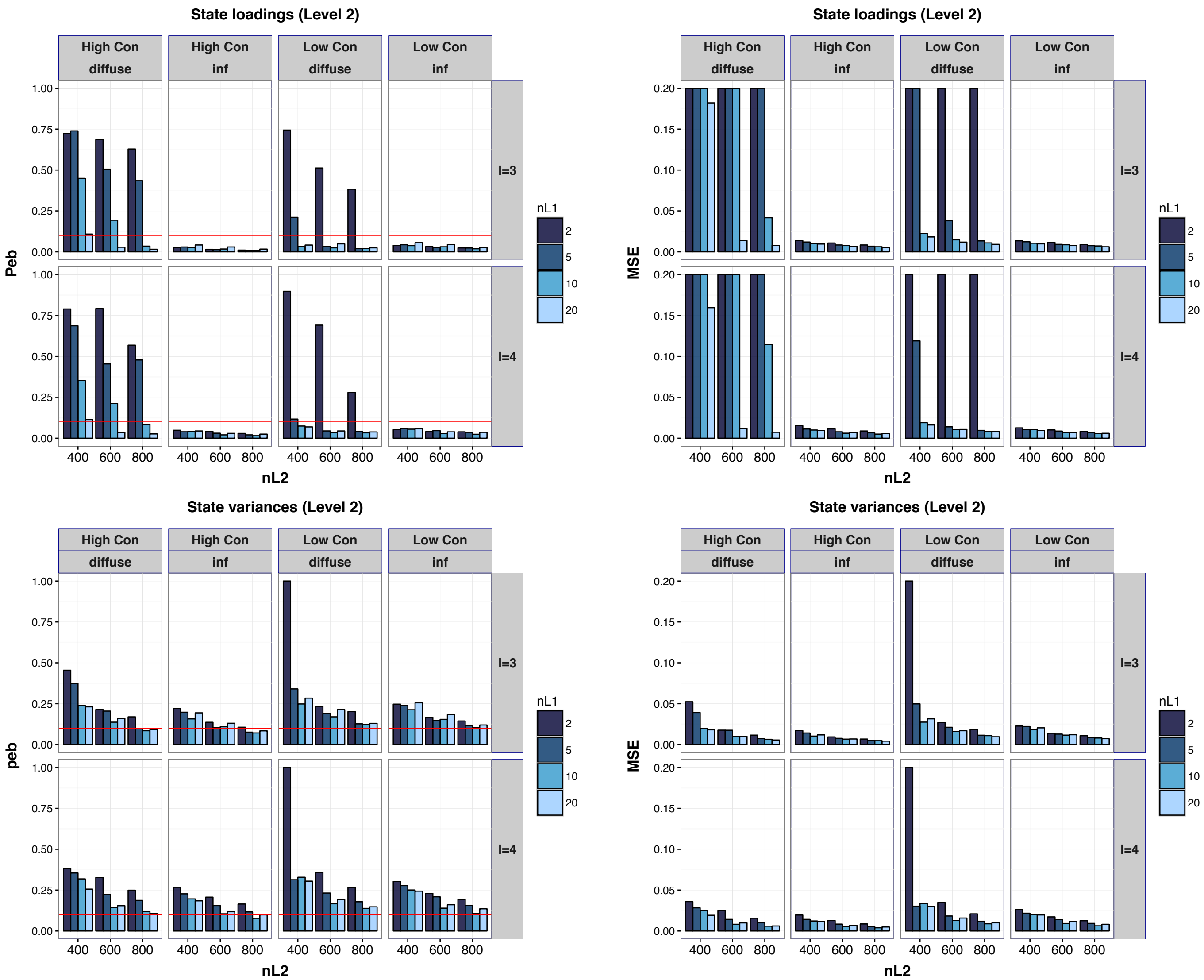

Figure C 2: Parameter estimation bias (peb) and mean squared error (MSE) for latent state residual factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

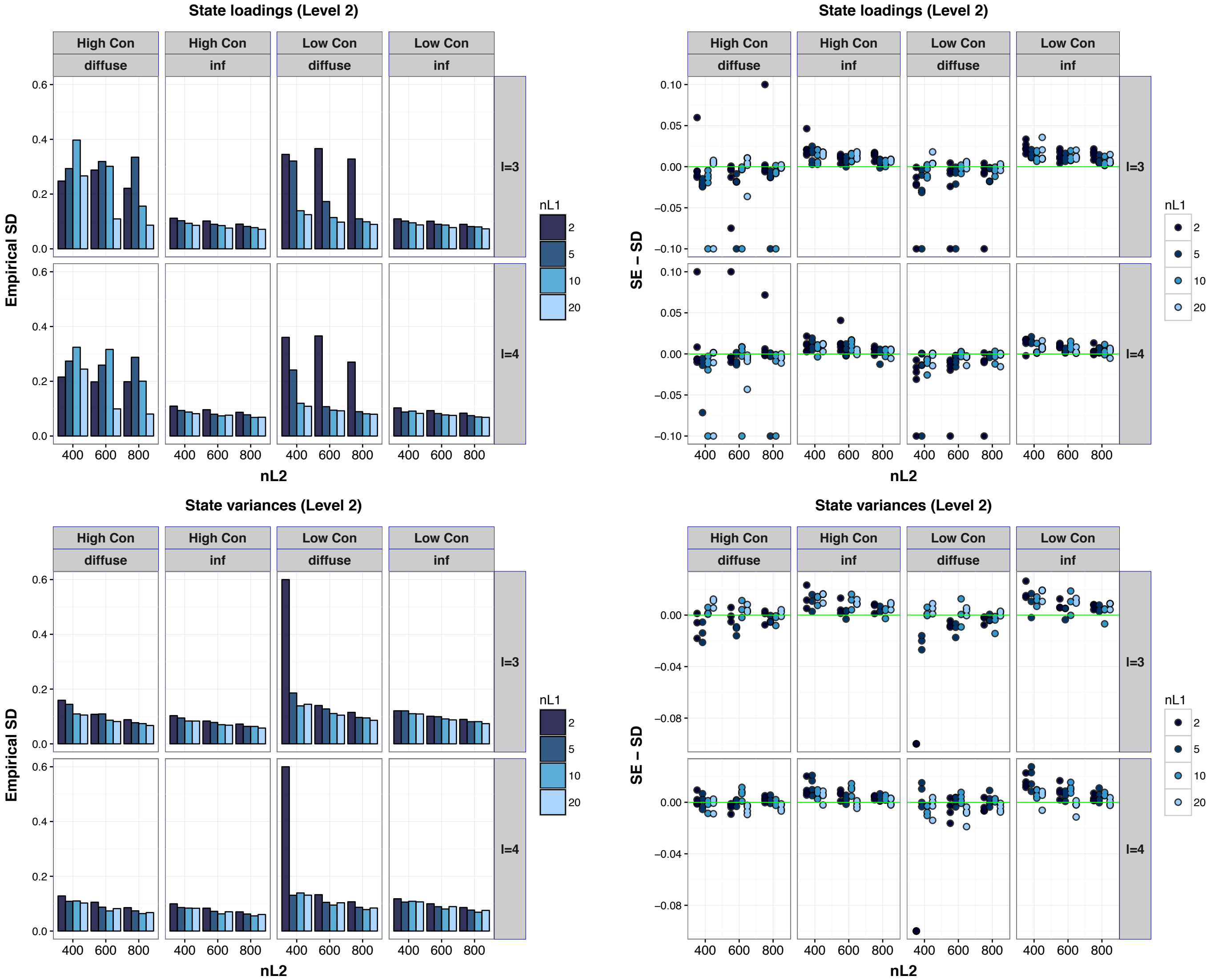

Figure C 3: Empirical SDs and standard error bias (SE - SD) for latent state residual factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

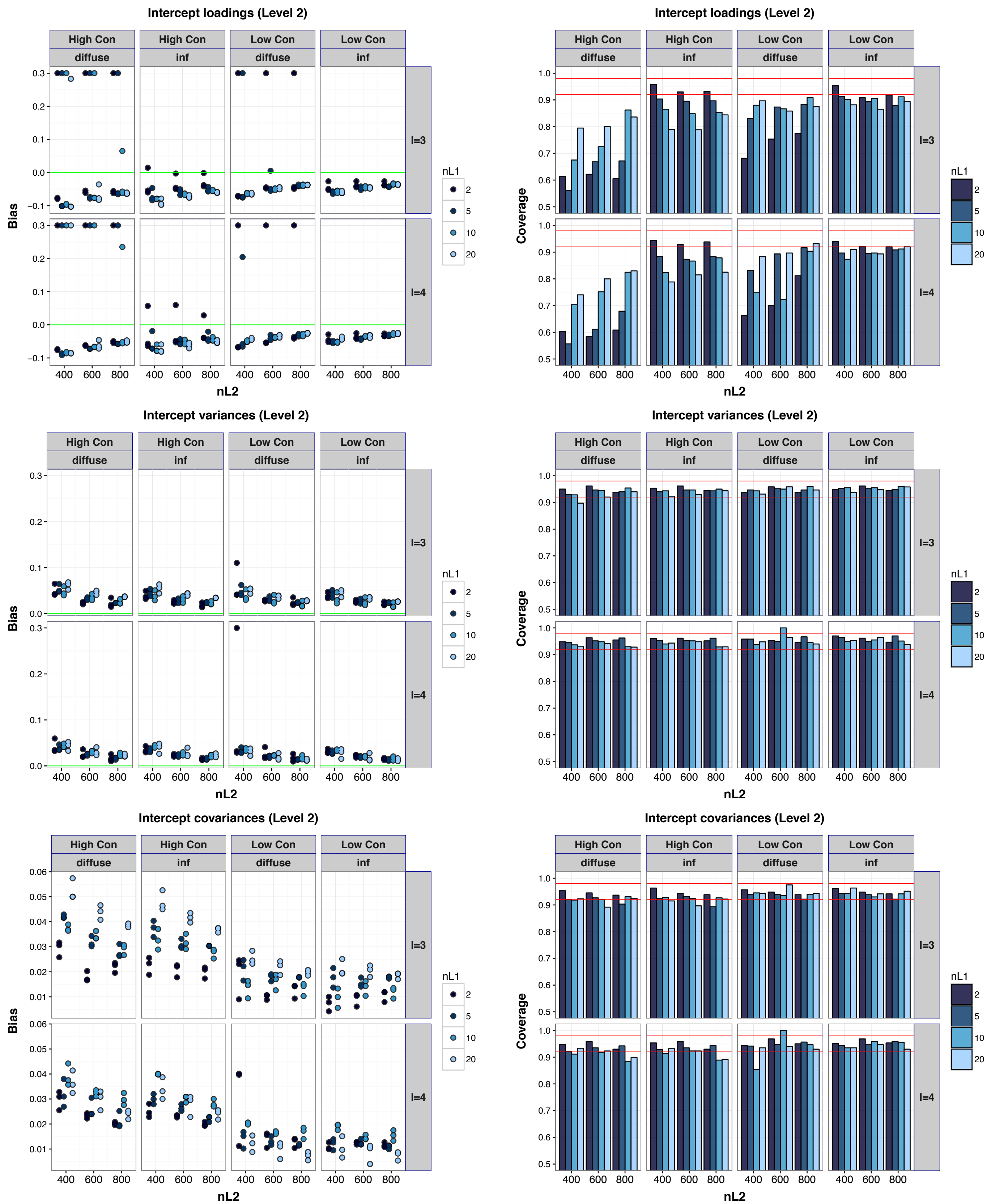

Figure C 4: Bias and 95% coverage for latent intercept factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values> 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

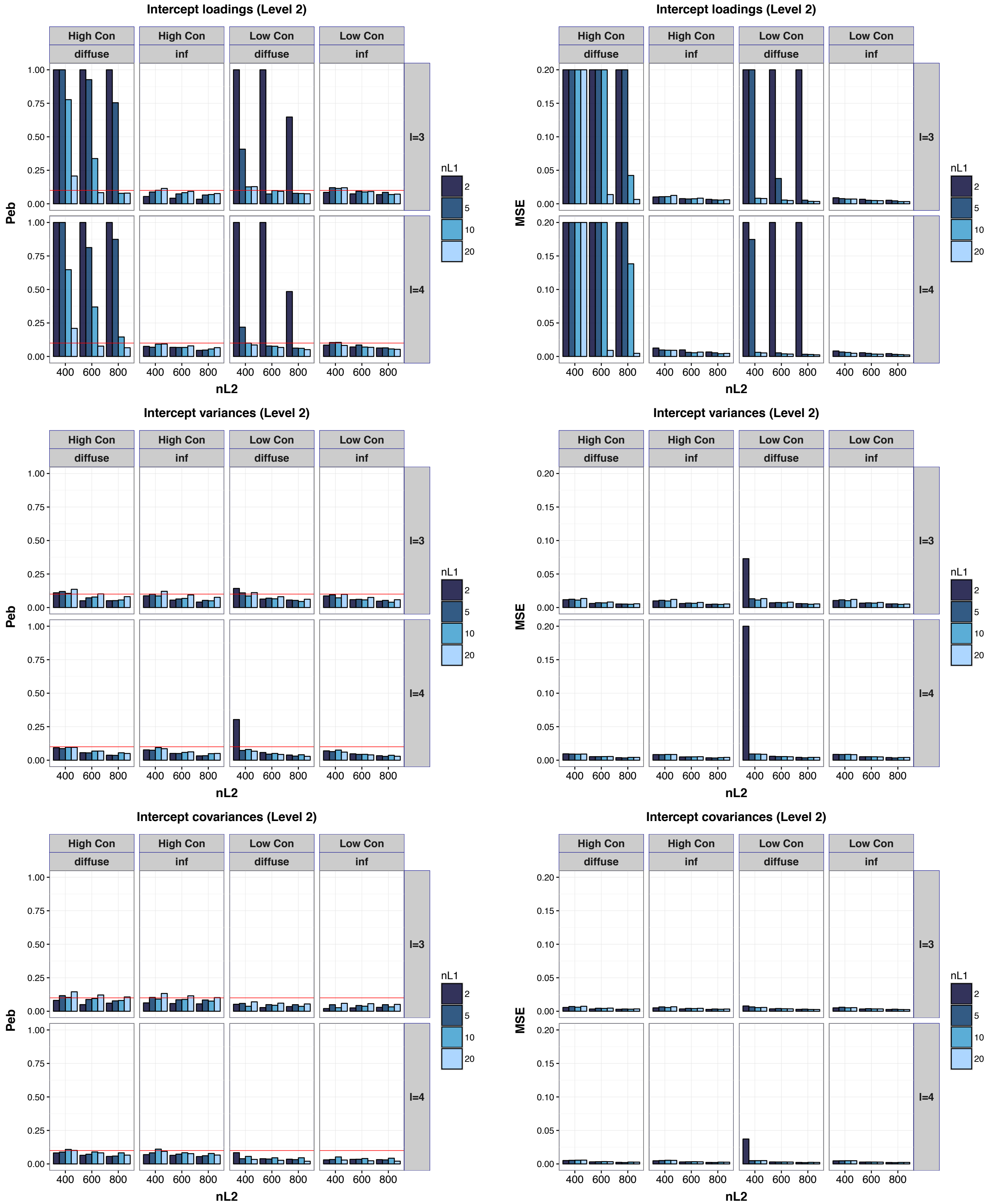

Figure C 5: Parameter estimation bias (peb) and mean squared error (MSE) for latent intercept factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

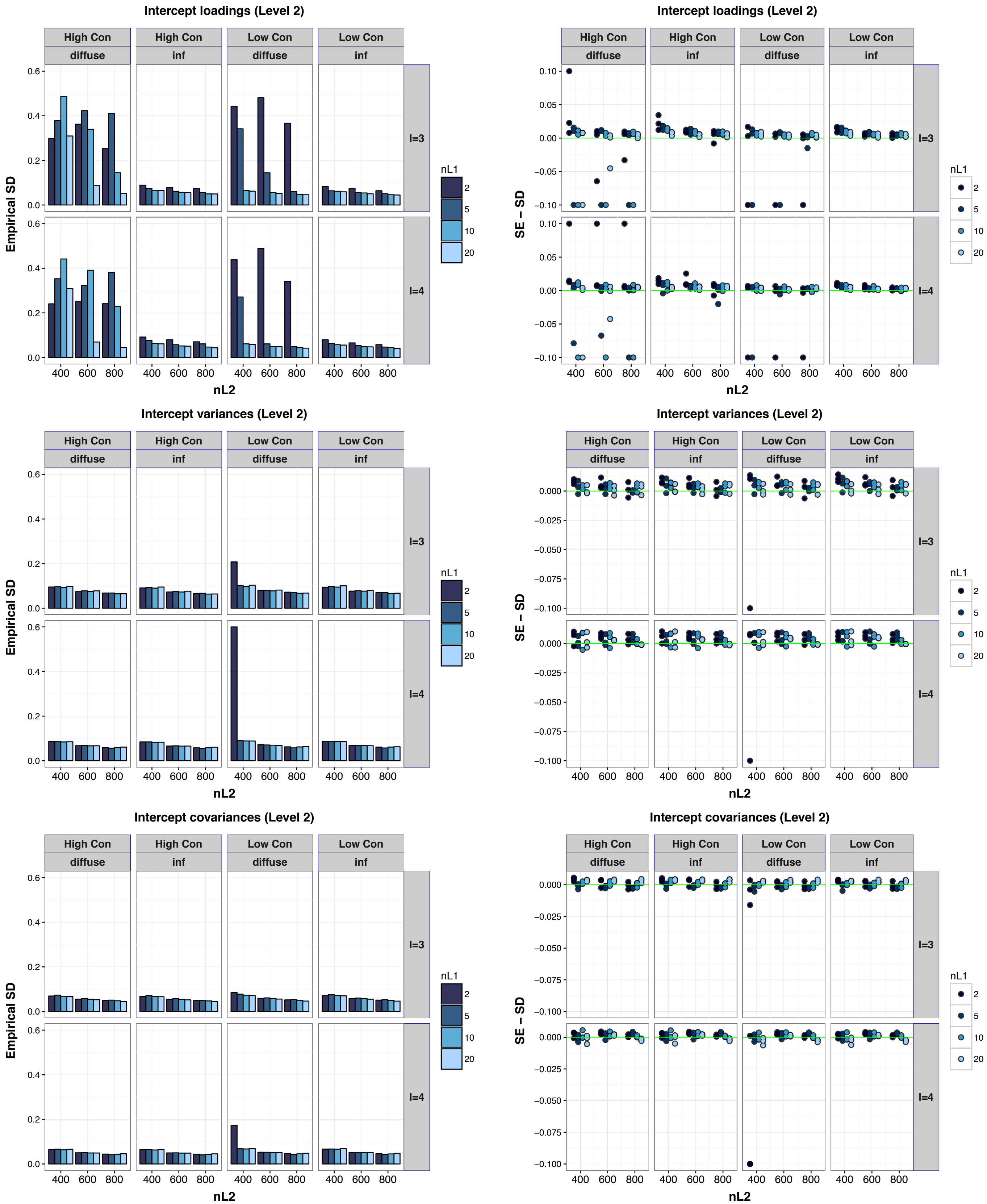

Figure C 6: Empirical SDs and standard error bias (SE - SD) for latent intercept factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

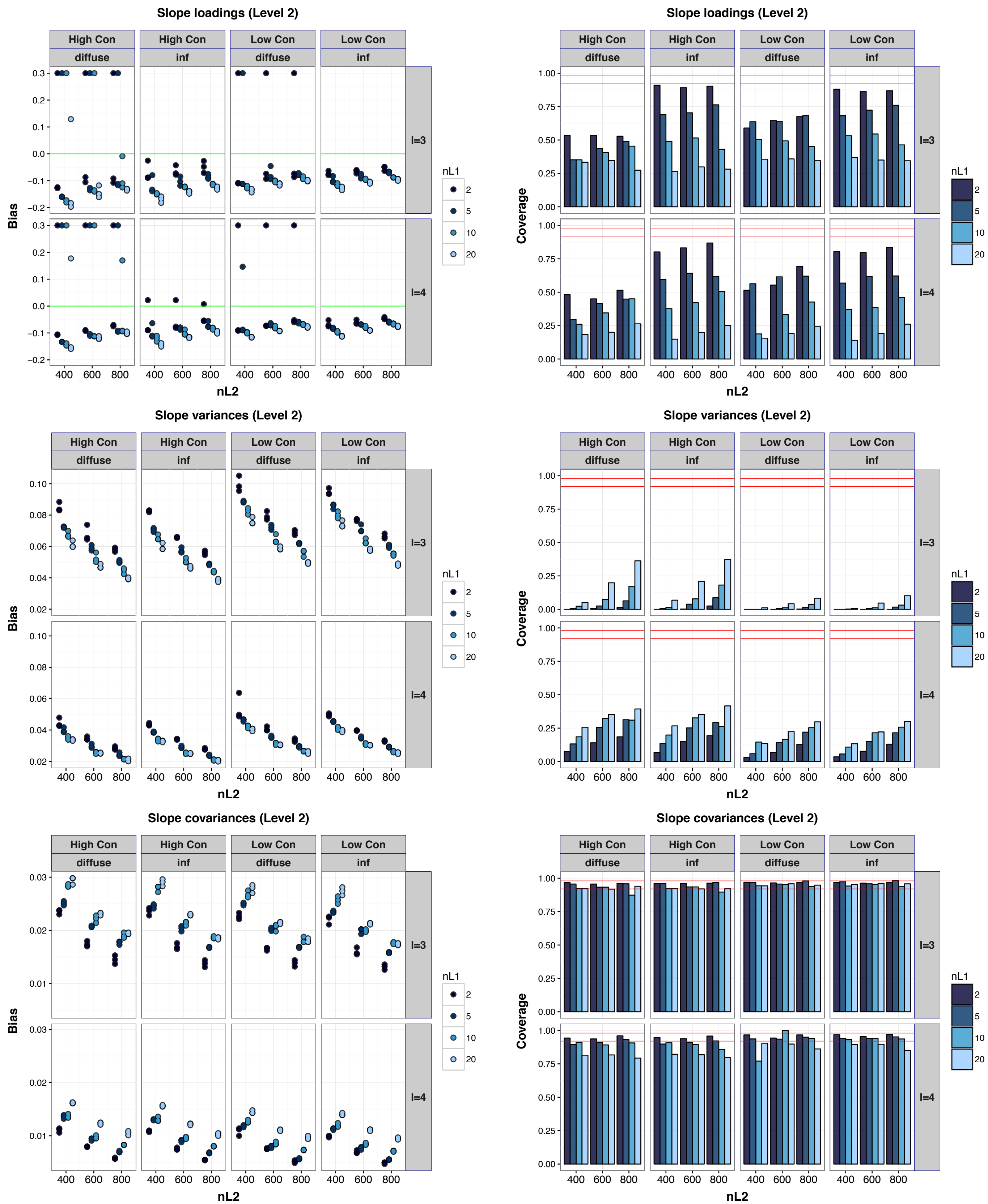

Figure C 7: Bias and 95% coverage for latent slope factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values> 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

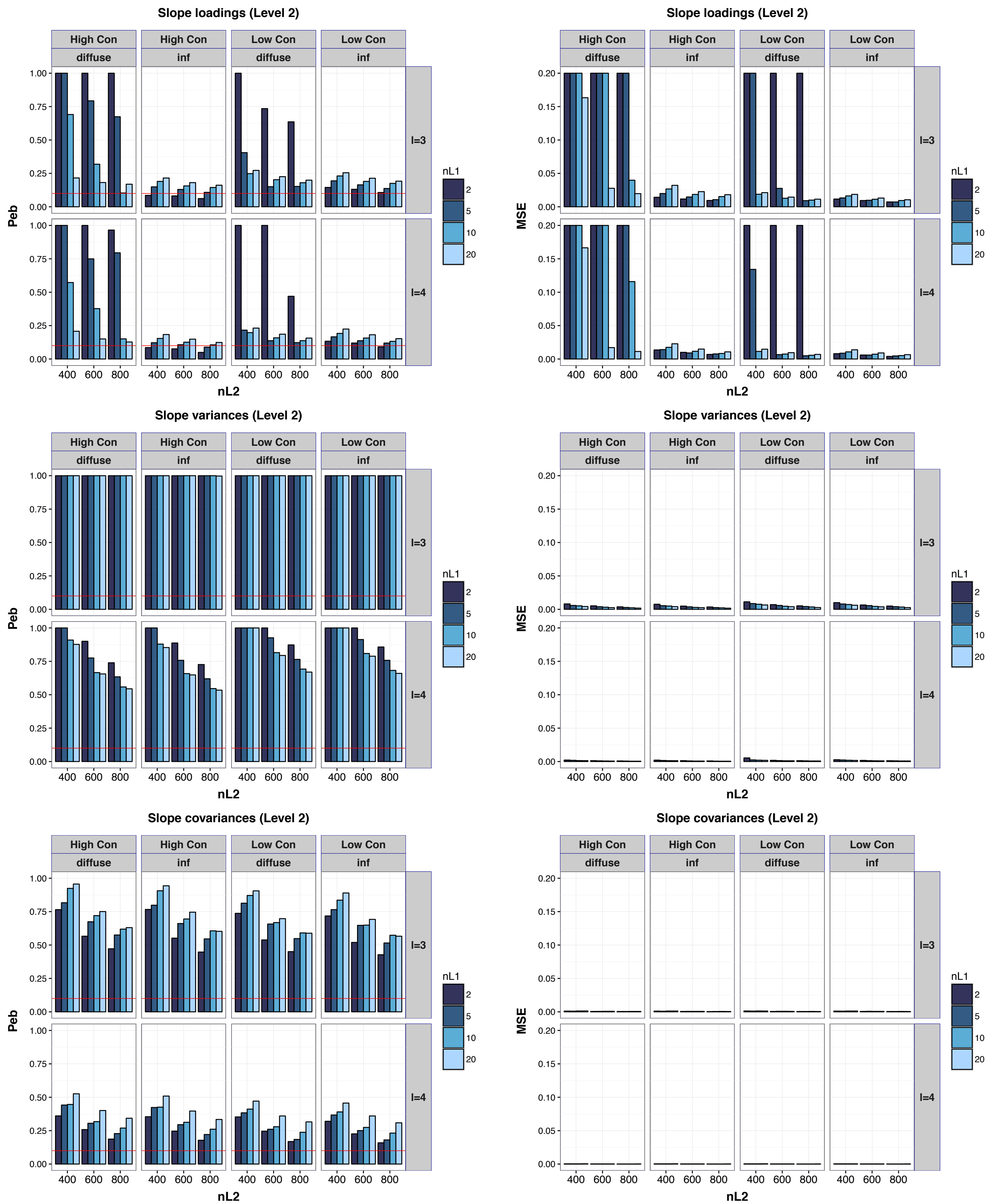

Figure C 8: Parameter estimation bias (peb) and mean squared error (MSE) for latent slope factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

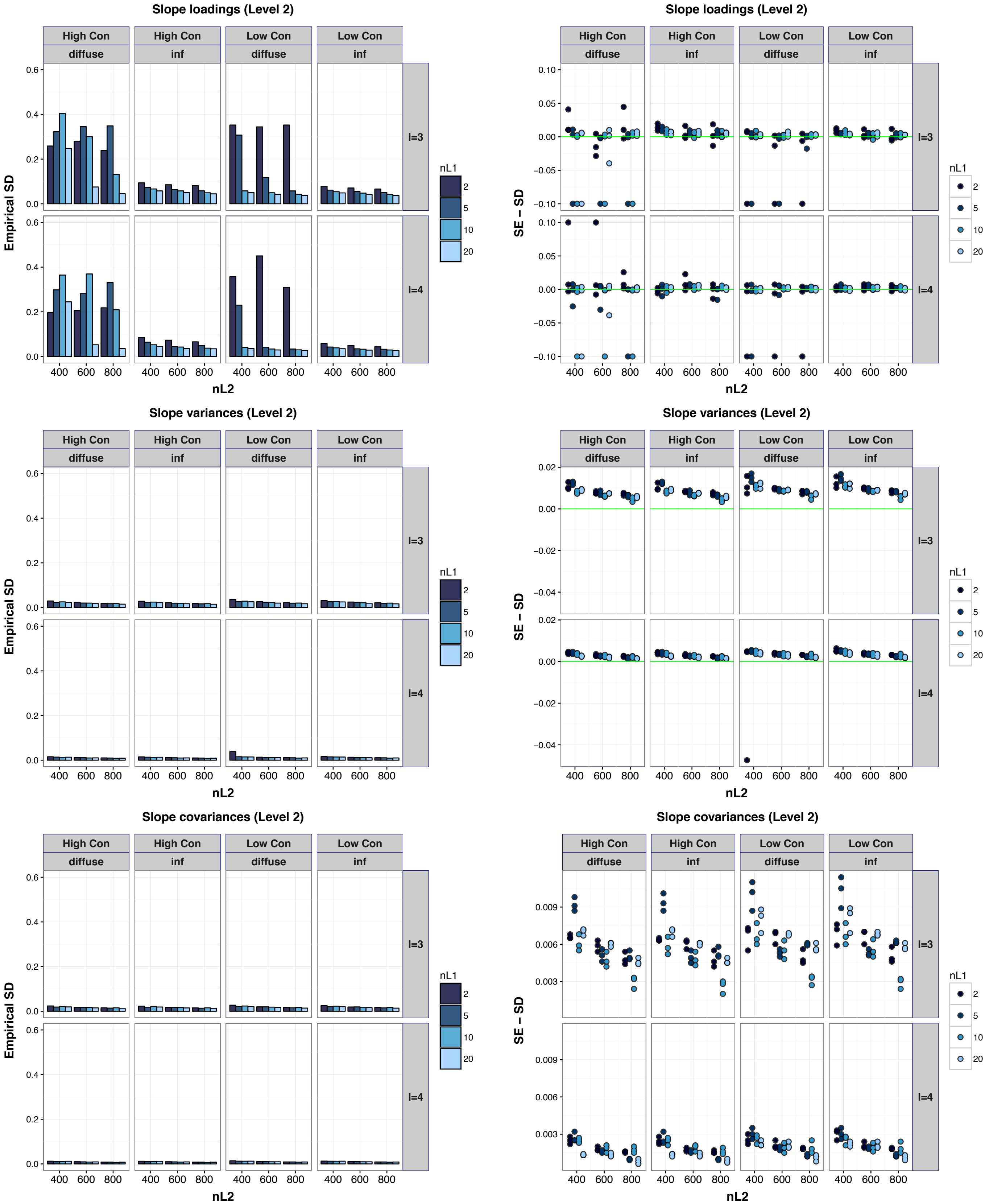

Figure C 9: Empirical SDs and standard error bias (SE - SD) for latent slope factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

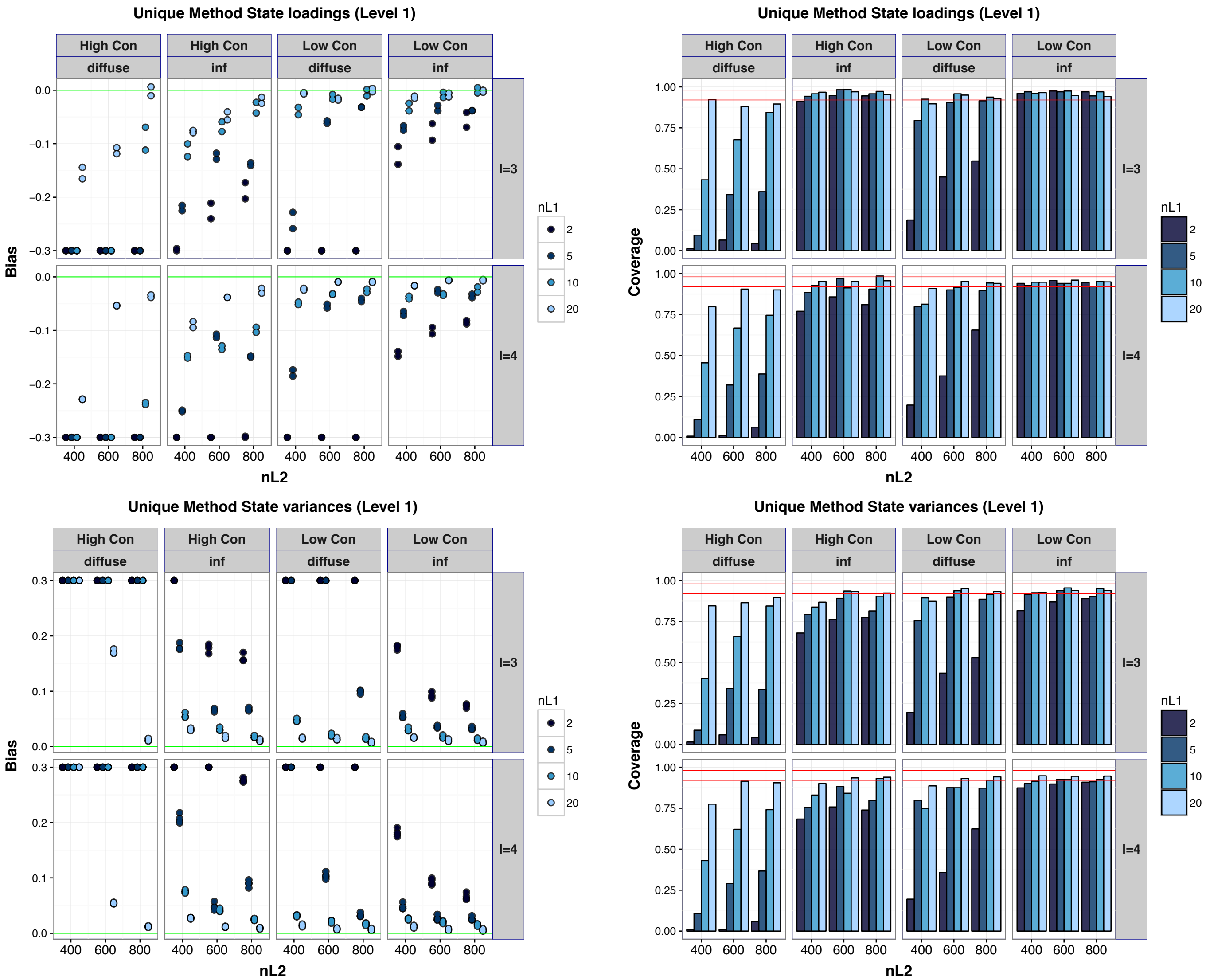

Figure C 10: Bias and 95% coverage for unique method state residual factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

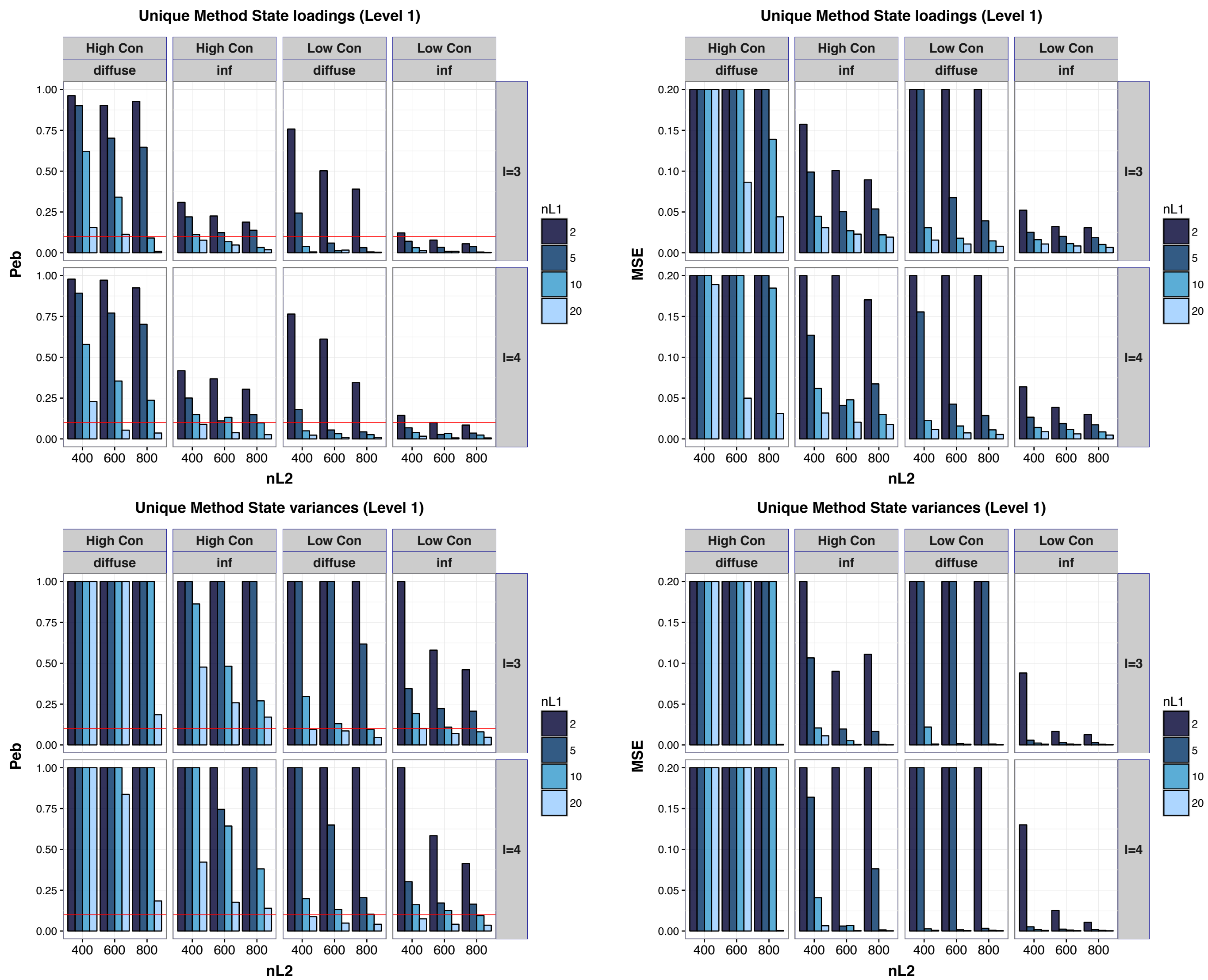

Figure C 11: Parameter estimation bias (peb) and mean squared error (MSE) for unique method state residual factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

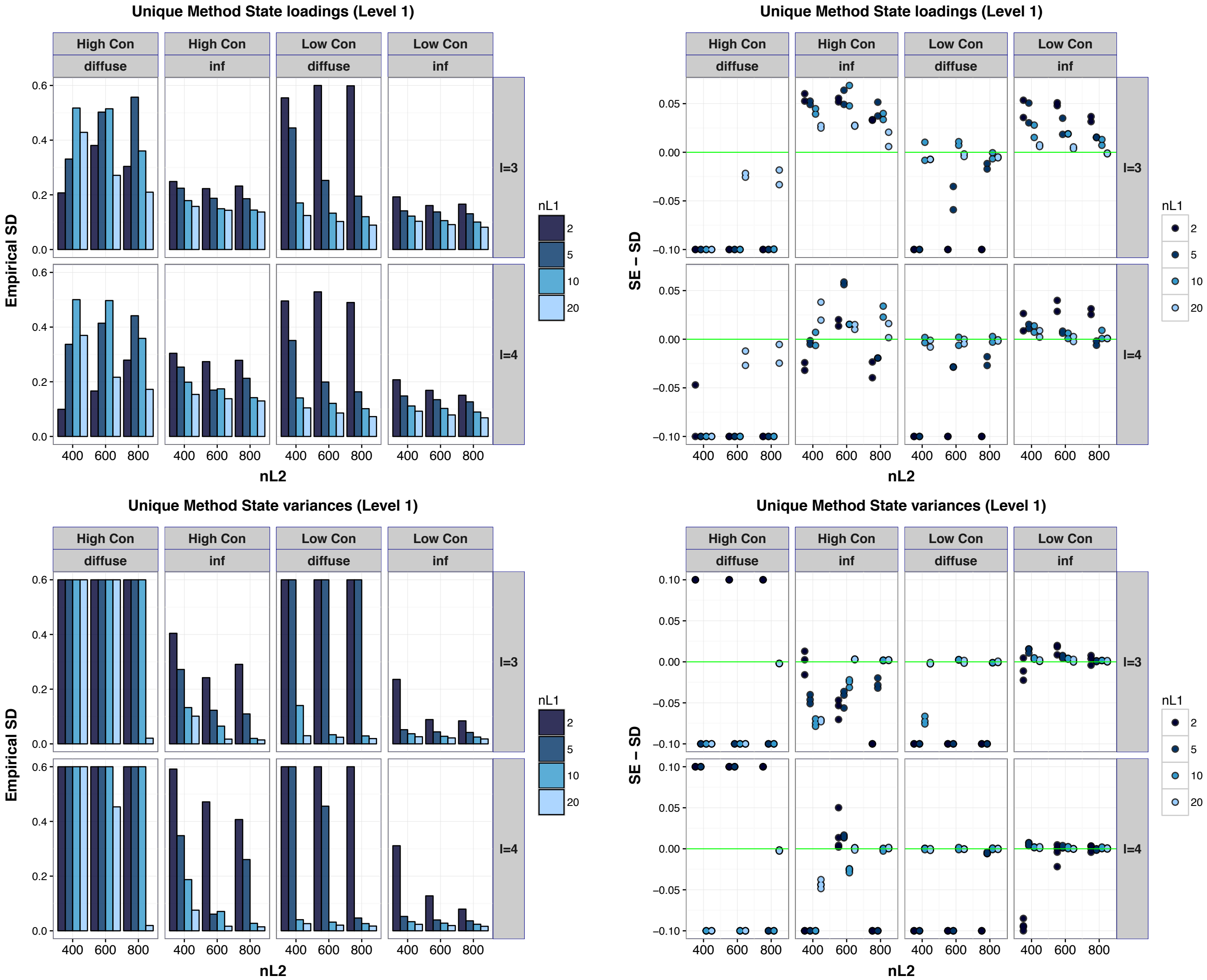

Figure C 12: Empirical SDs and standard error bias (SE - SD) for unique method state residual factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

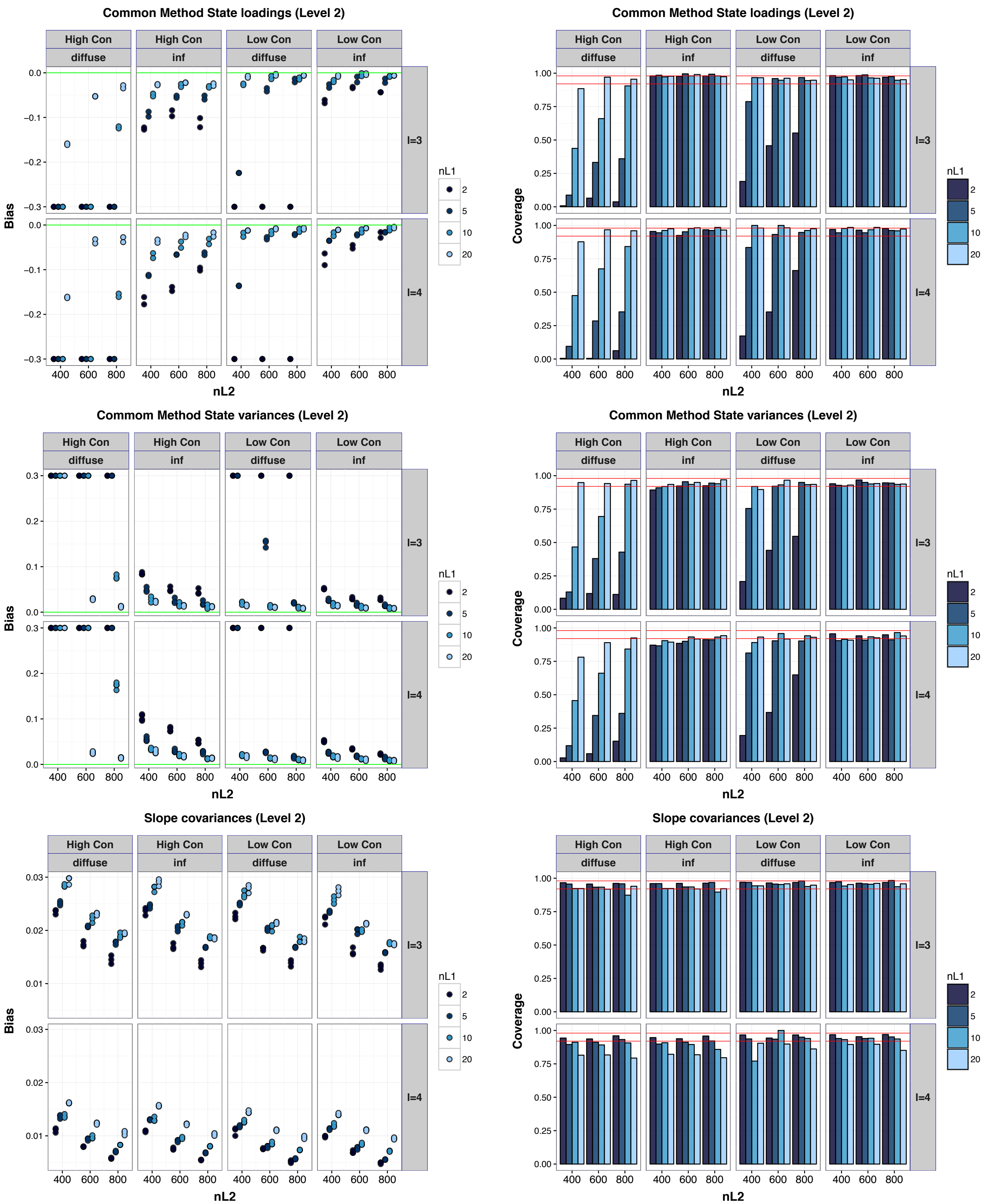

Figure C 13: Bias and 95% coverage for common method state residual factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

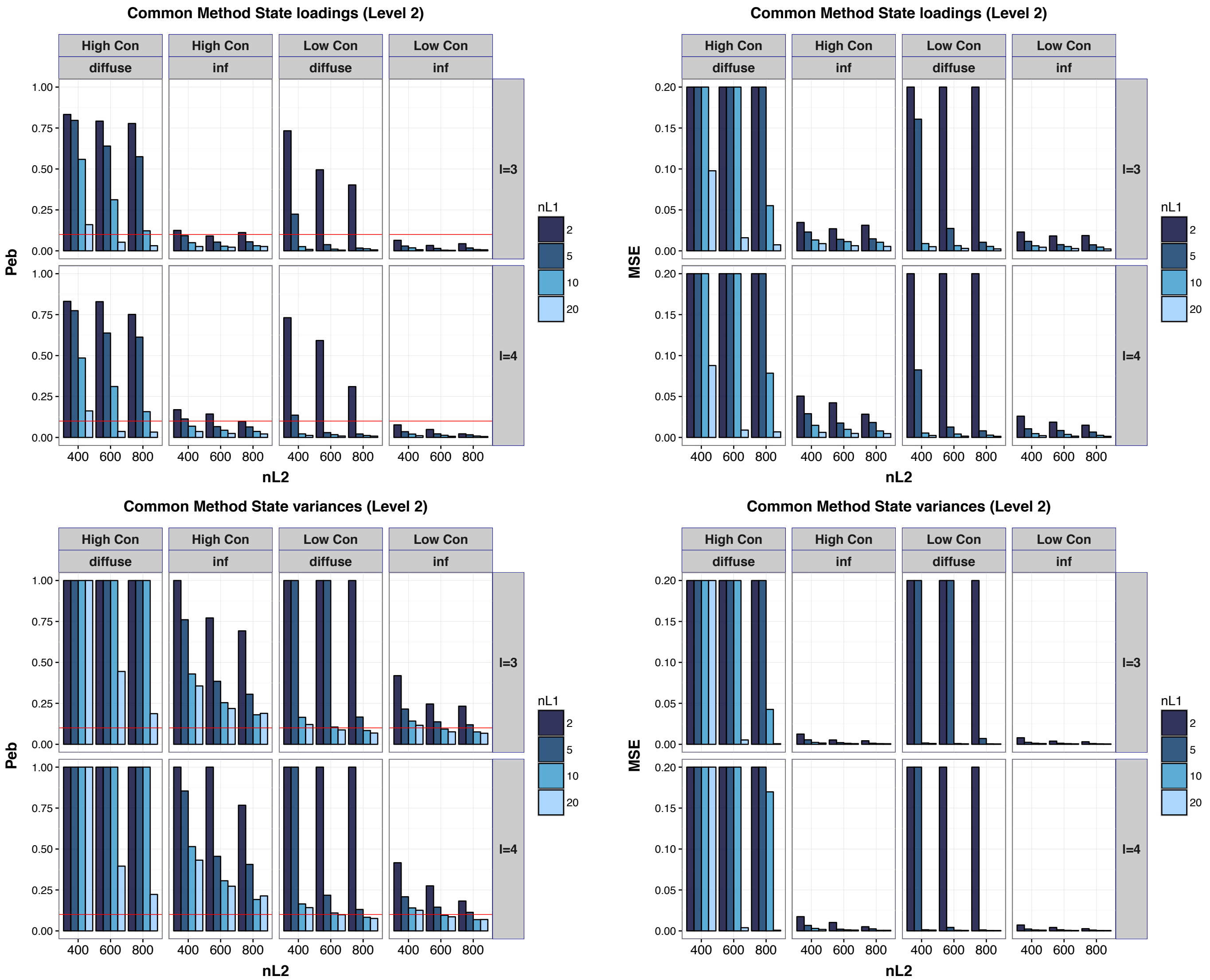

Figure C 14: Parameter estimation bias (peb) and mean squared error (MSE) for common method state residual factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

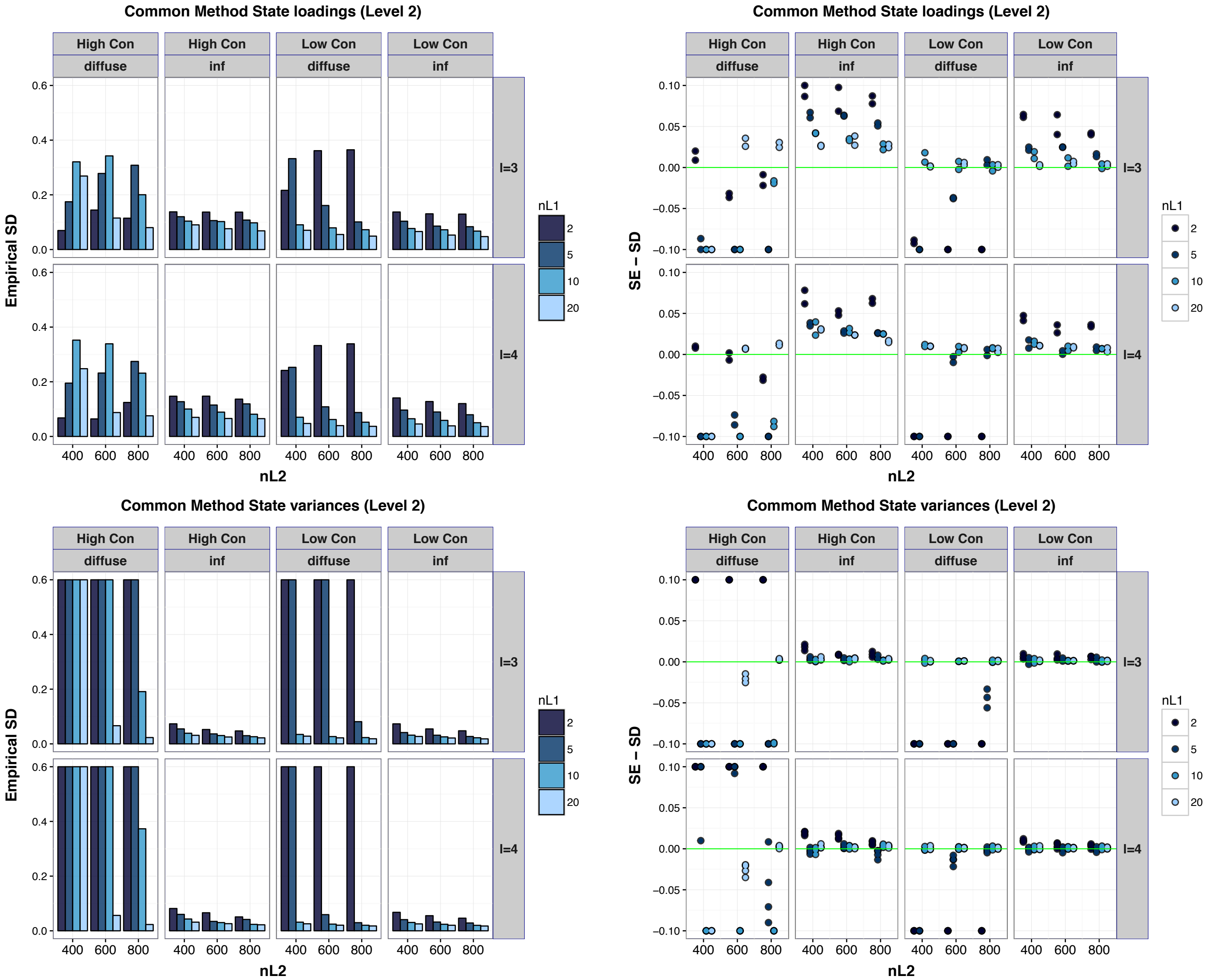

Figure C 15: Empirical SDs and standard error bias (SE - SD) for common method state residual factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

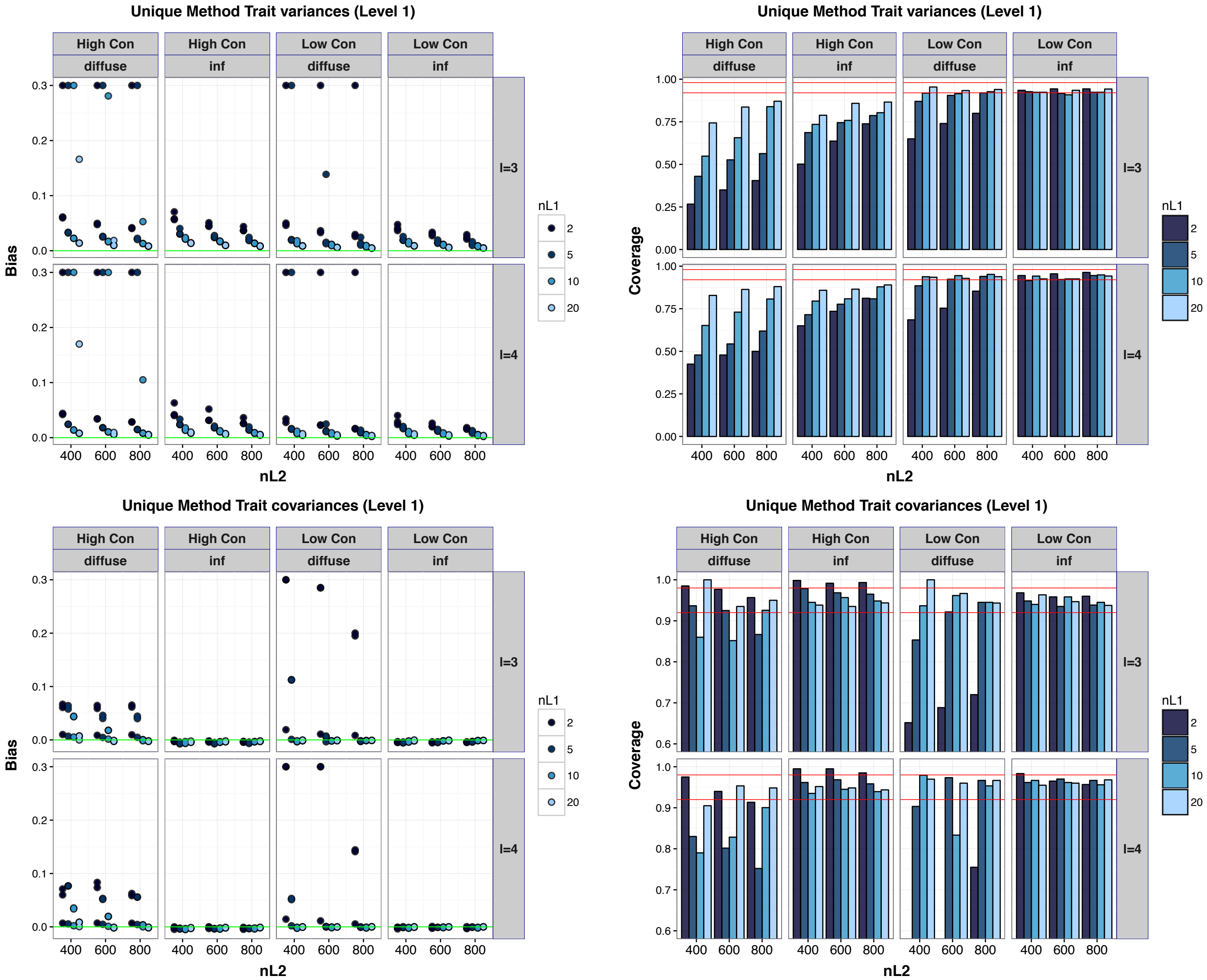

Figure C 16: Bias and 95% coverage for unique method trait factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

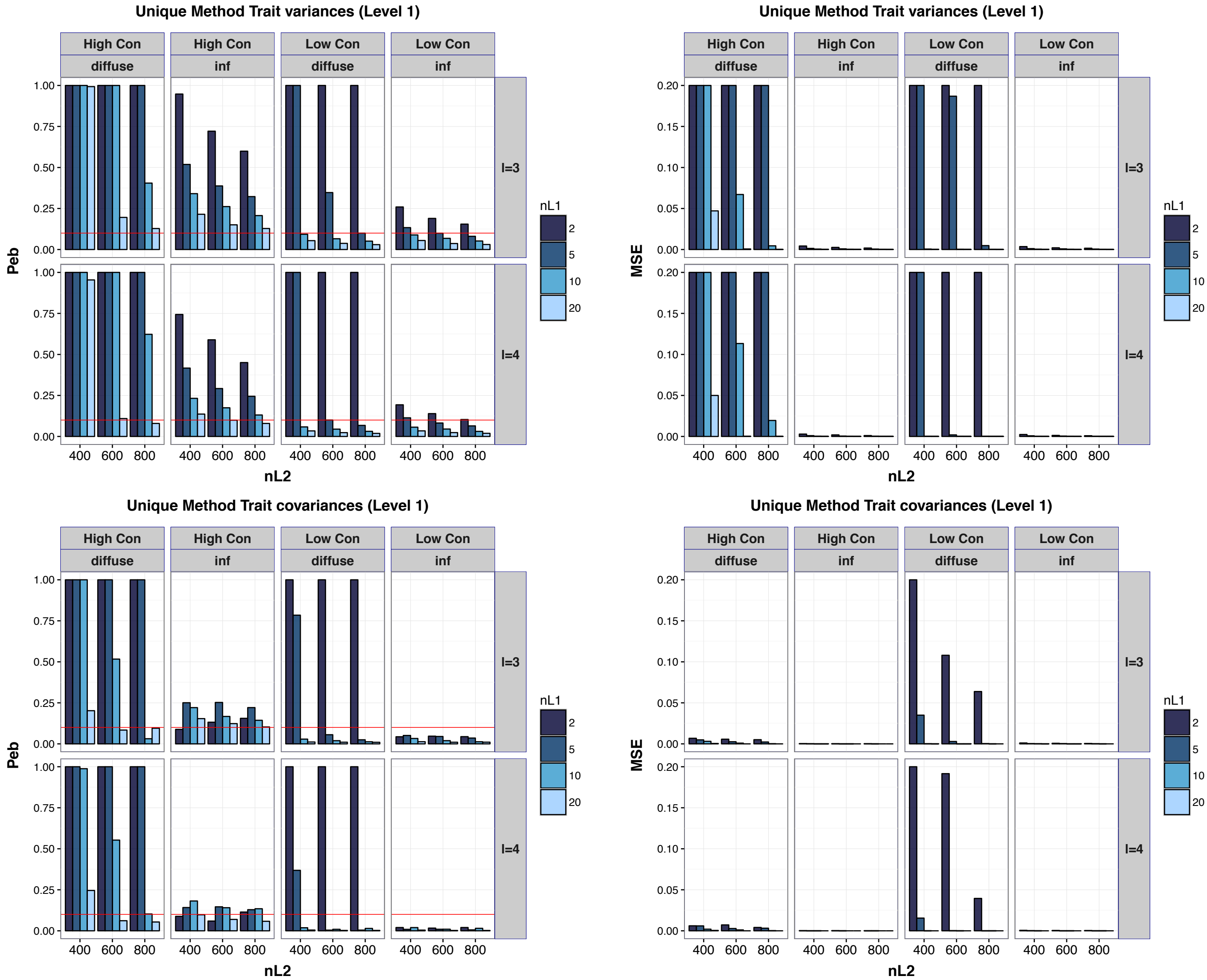

Figure C 17: Parameter estimation bias (peb) and mean squared error (MSE) for unique method trait factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

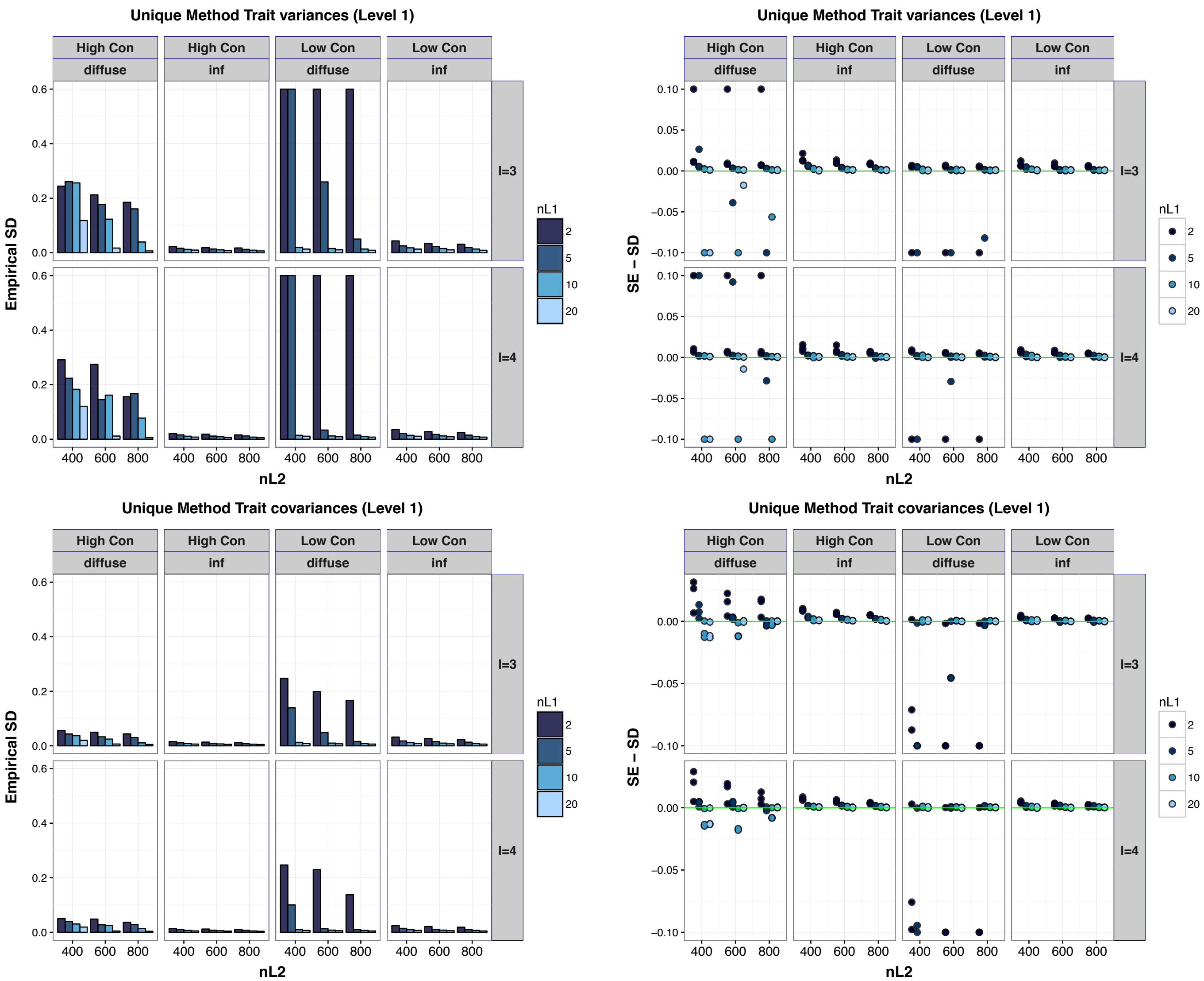

Figure C 18: Empirical SDs and standard error bias (SE - SD) for unique method trait factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

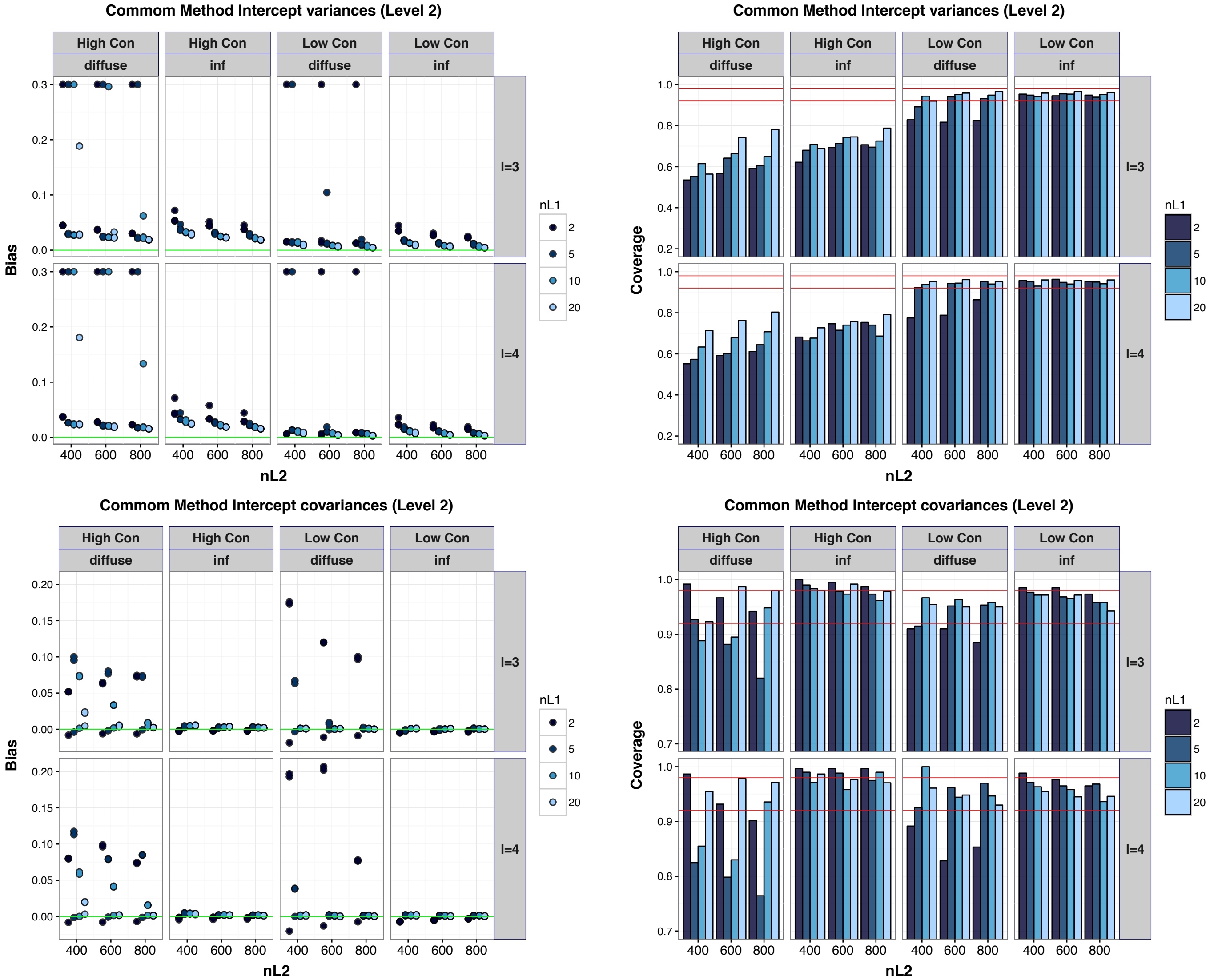

Figure C 19: Bias and 95% coverage for common method intercept factors in the LGC-Com GRM with one construct. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

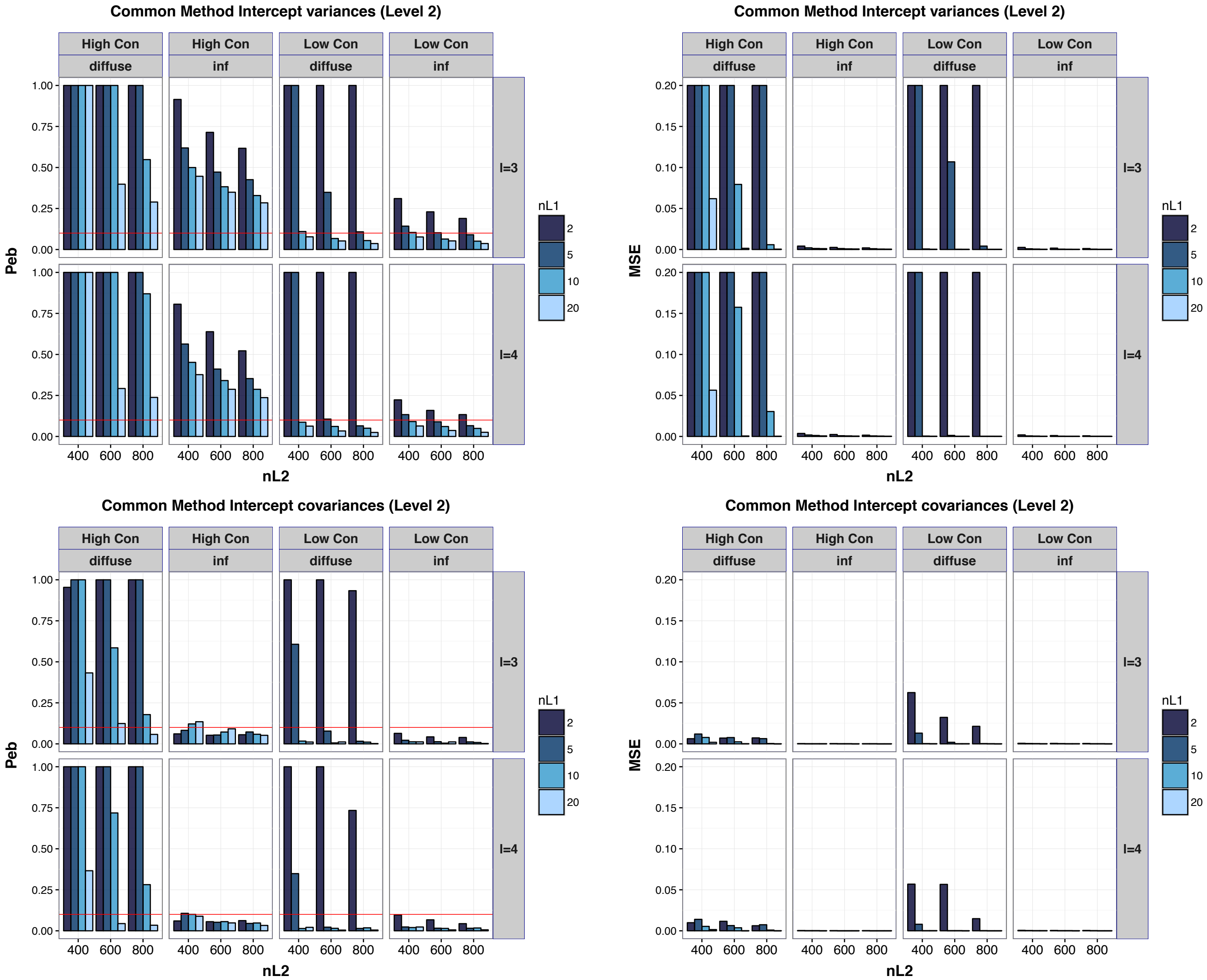

Figure C 20: Parameter estimation bias (peb) and mean squared error (MSE) for common method intercept factors in the LGC-Com GRM with one construct. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

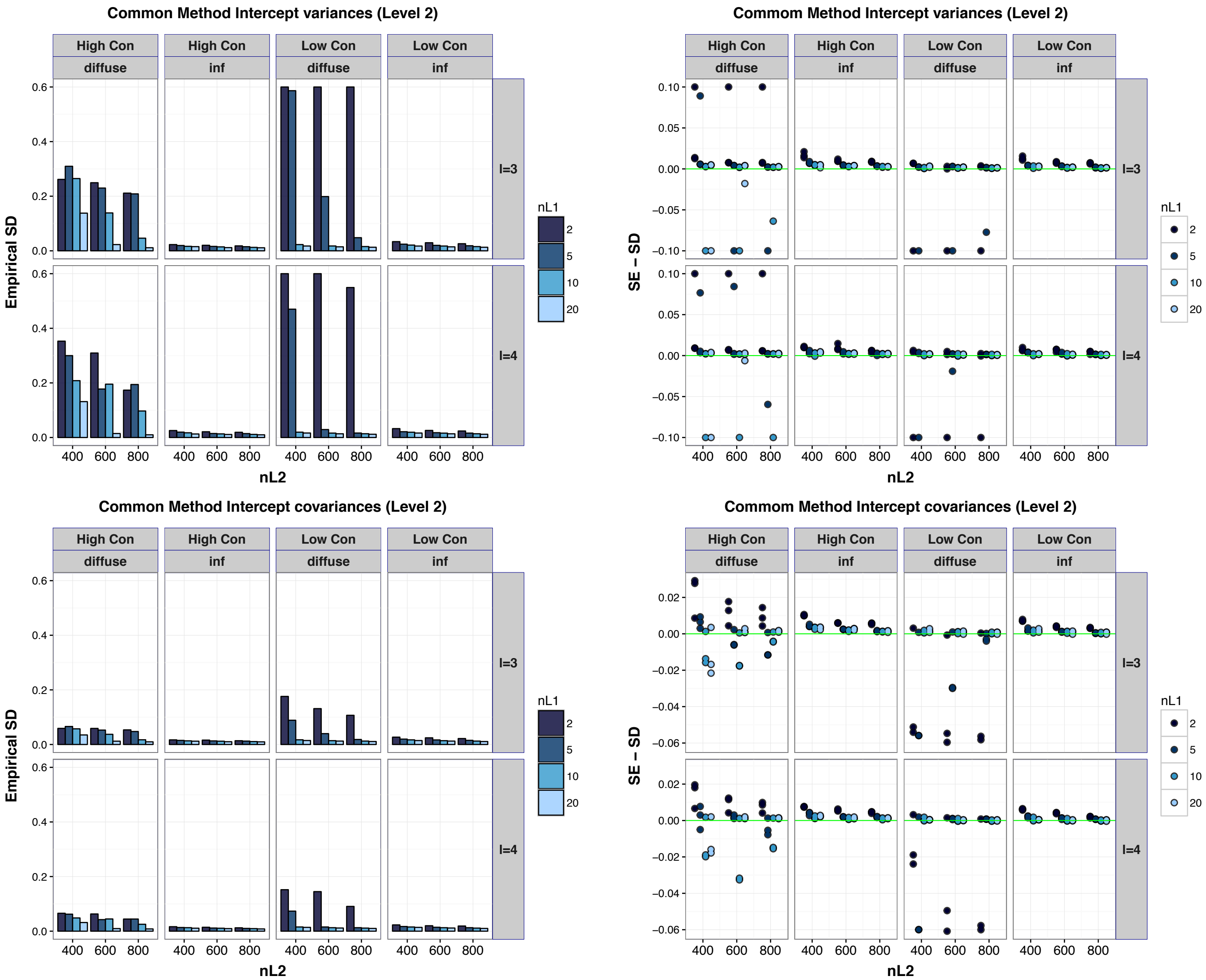

Figure C 21: Empirical SDs and standard error bias (SE - SD) for common method intercept factors in the LGC-Com GRM with one construct. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences $>$ 0.1 or $<$ -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

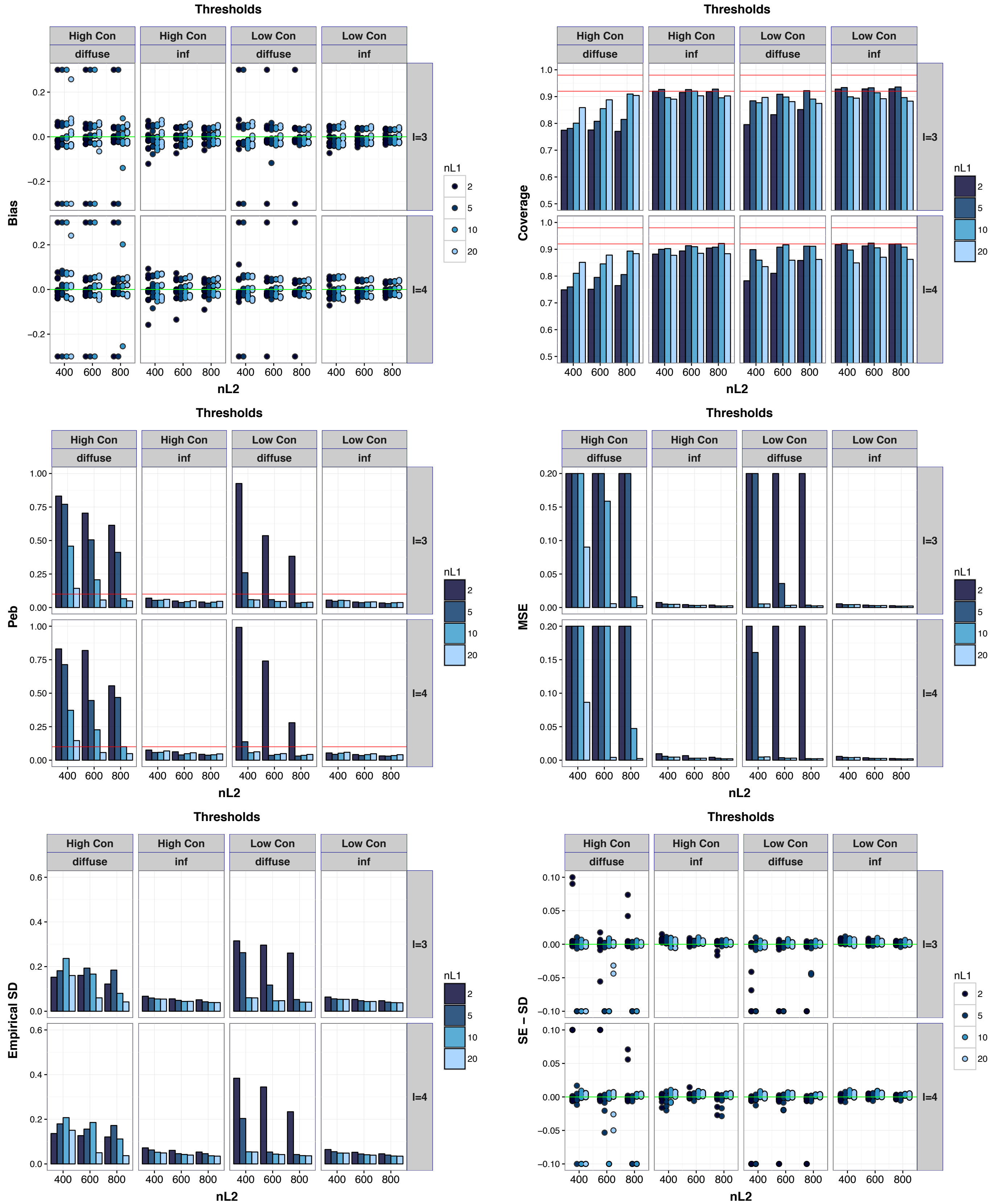

Figure C 22: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LGC-Com GRM with one construct. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2, Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

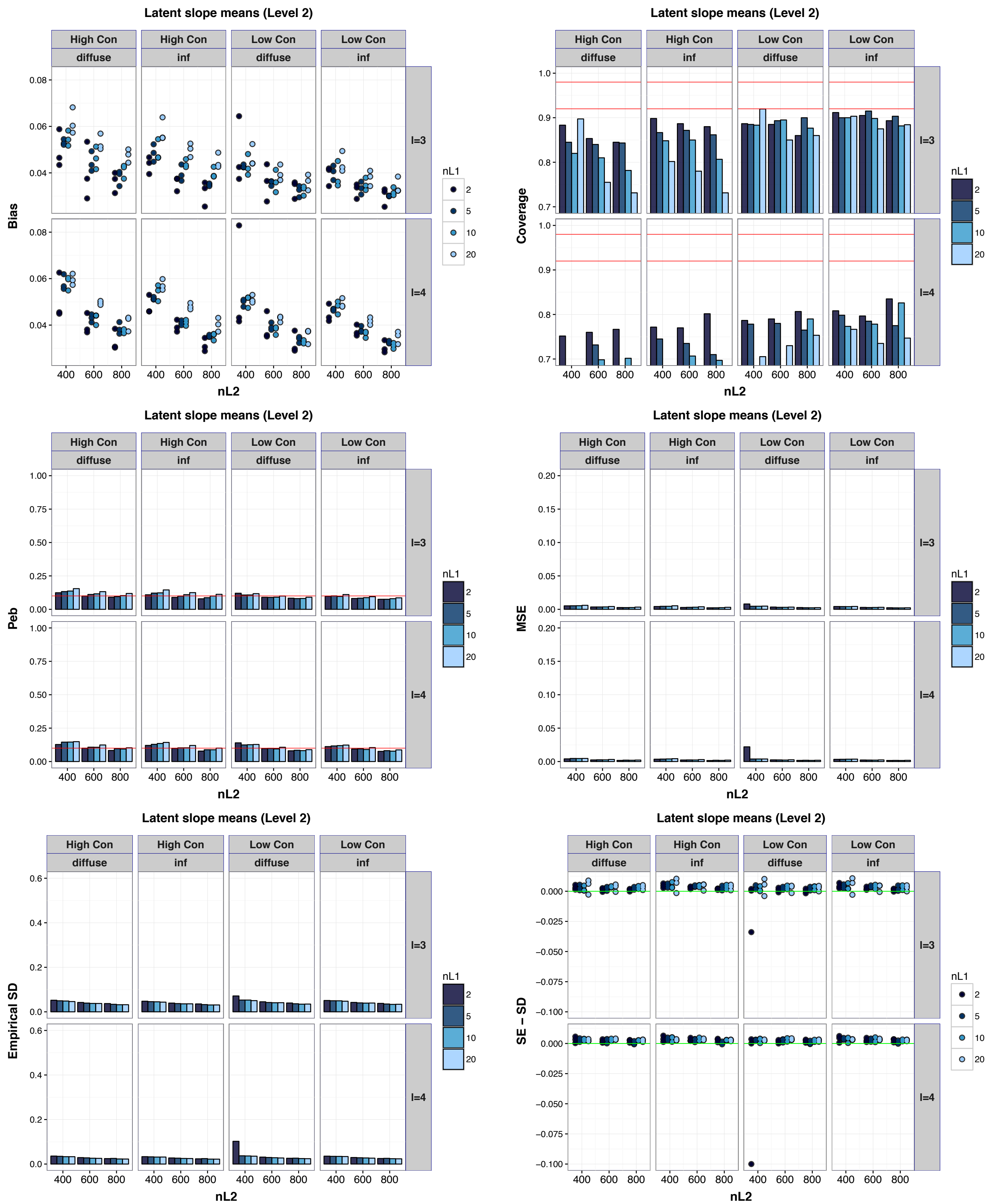

Figure C 23: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and Standard error bias (SE - SD) for the latent slope means in the LGC-Com GRM with one construct. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

C.3 Simulation results LGC-Com GRM. Case of two constructs ($j = 2$).

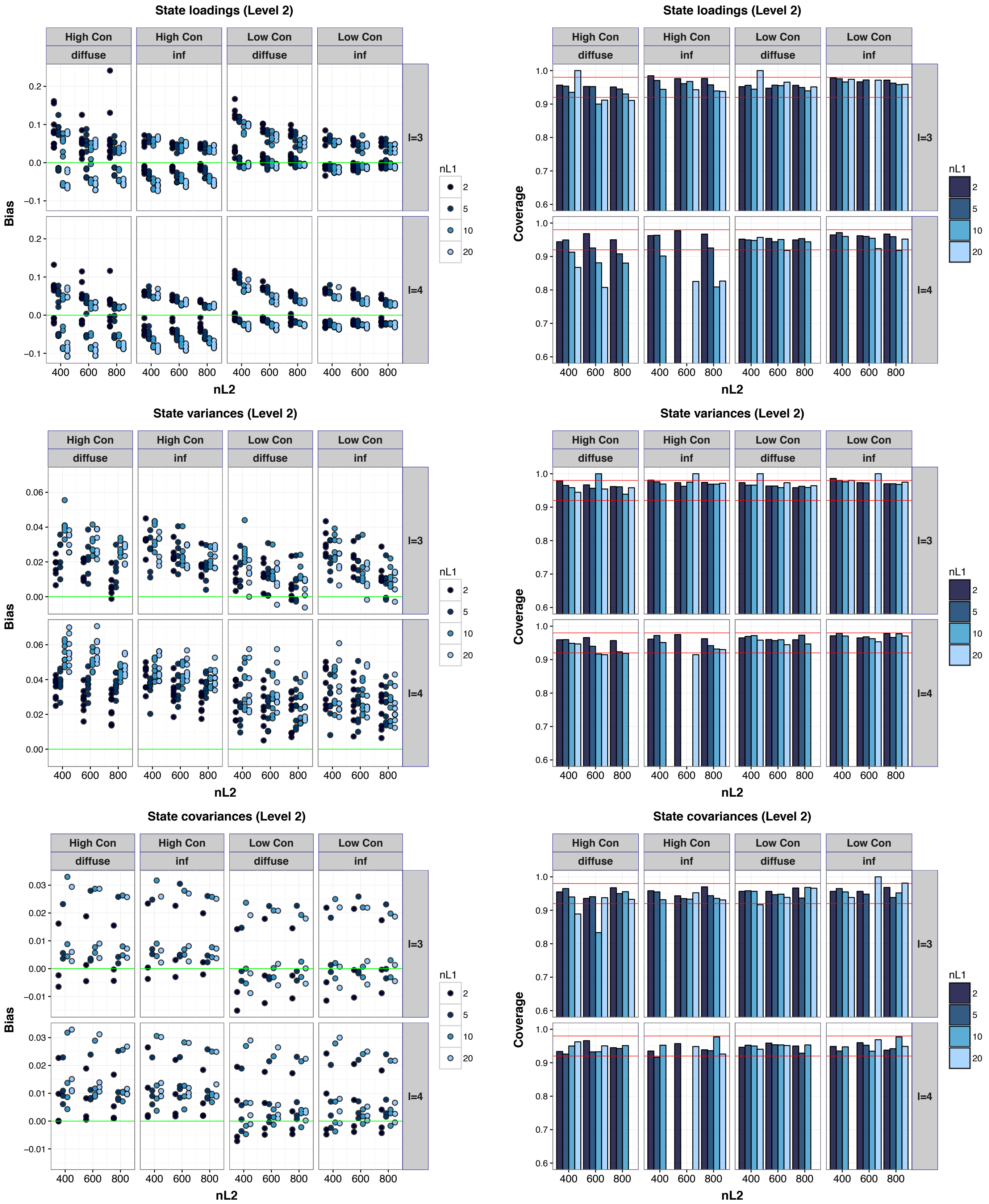

Figure C 24: Bias and 95% coverage for latent state residual factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

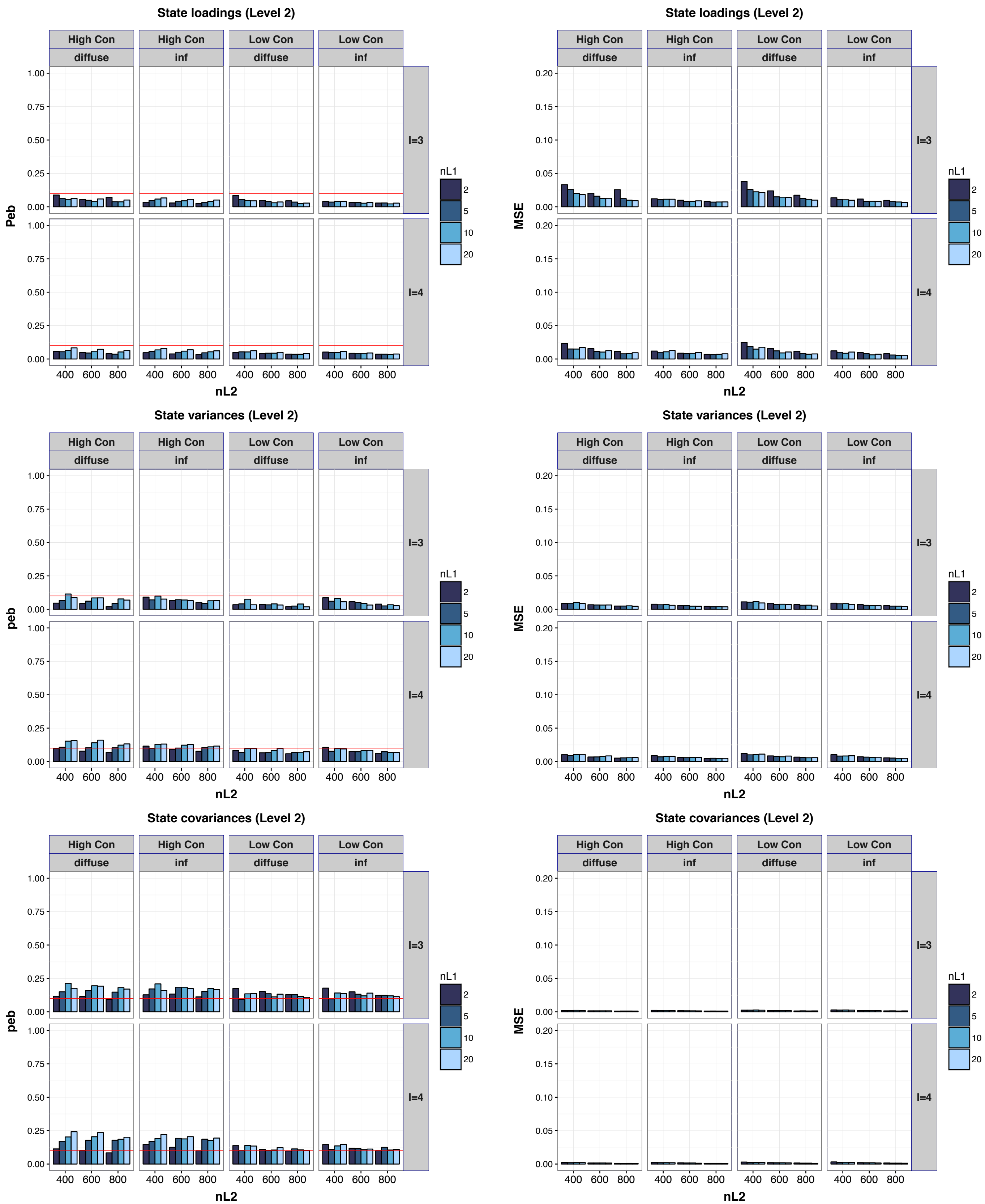

Figure C 25: Parameter estimation bias (peb) and mean squared error (MSE) for latent state residual factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

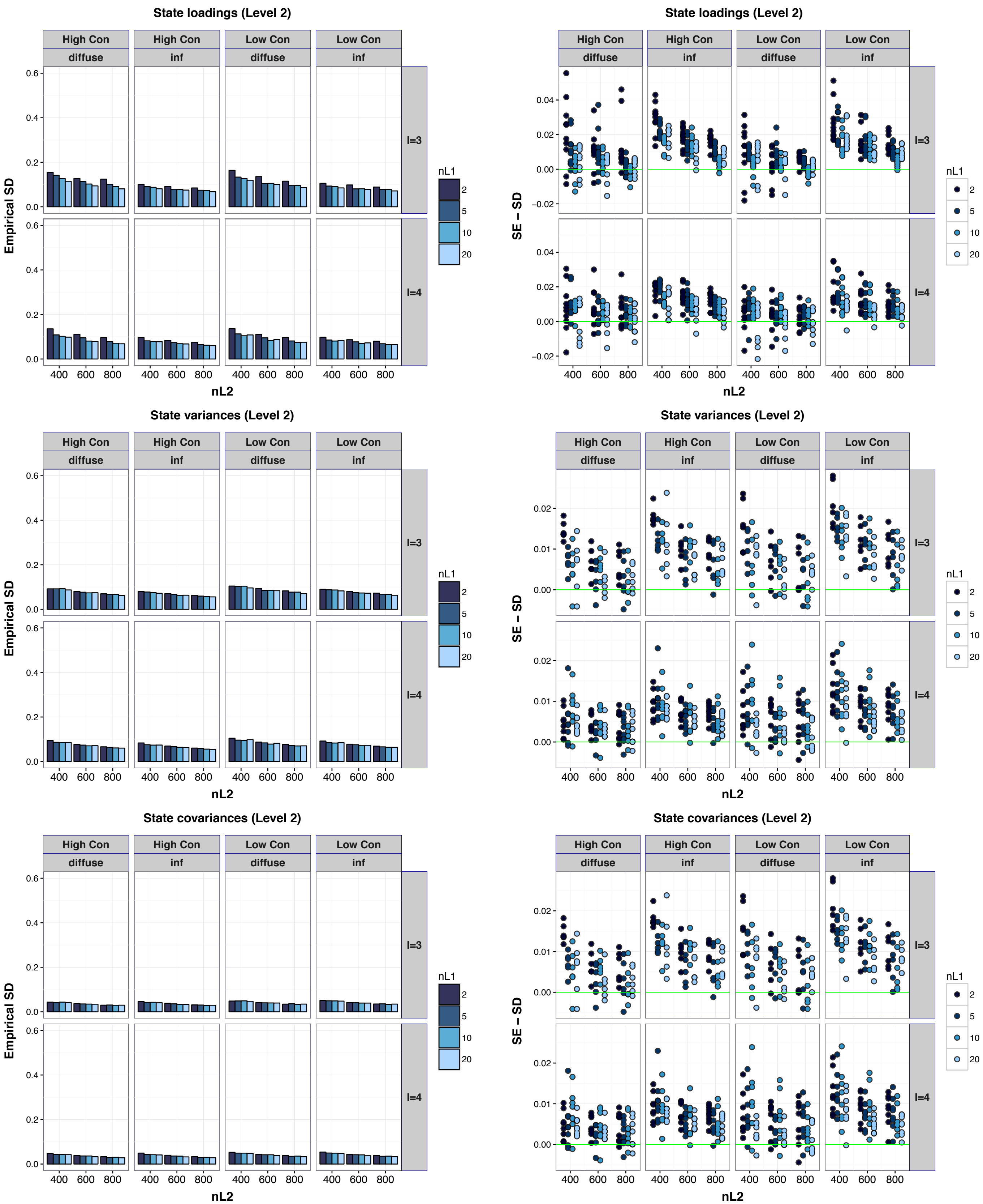

Figure C 26: Empirical SDs and standard error bias (SE - SD) for latent state residual factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

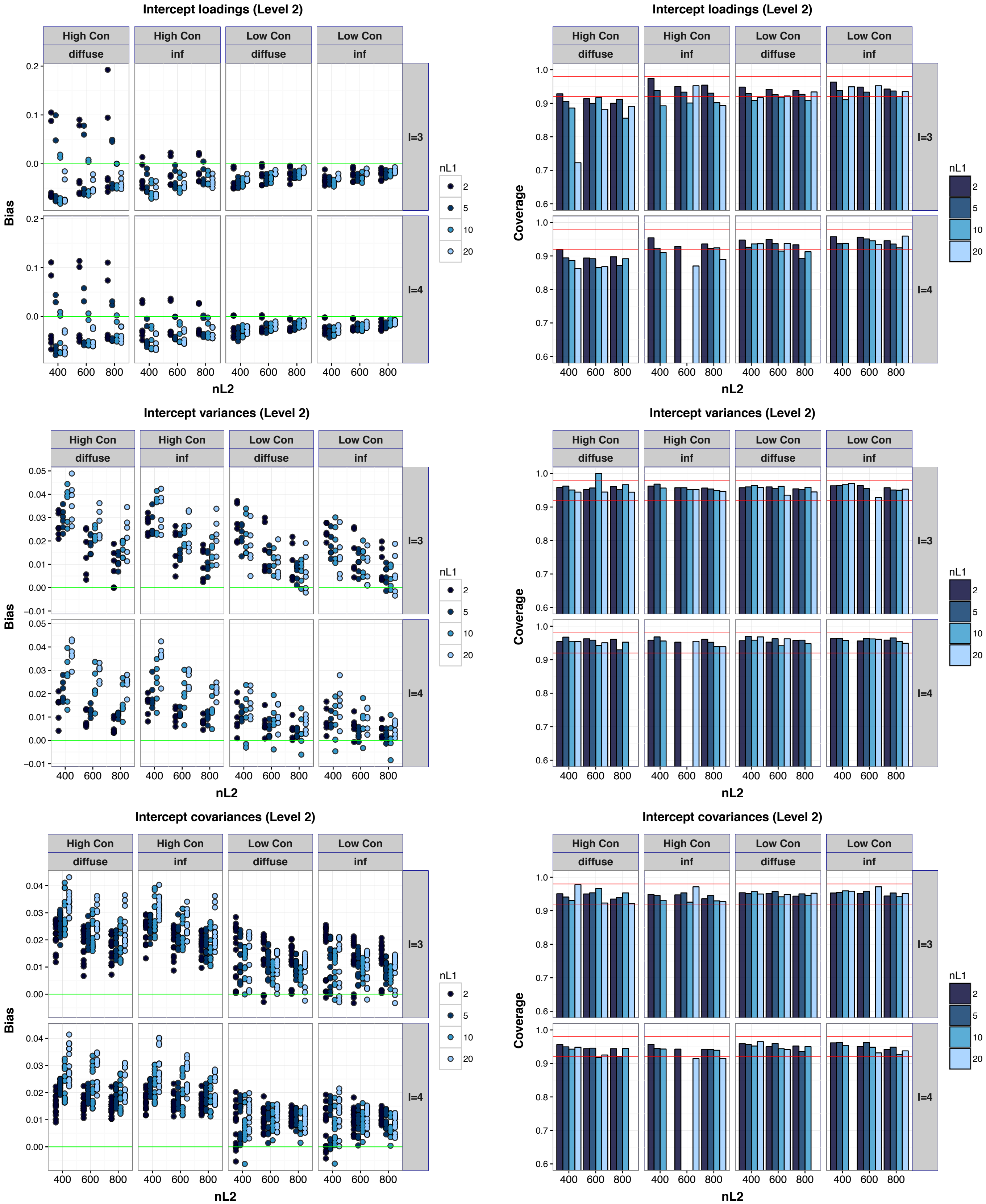

Figure C 27: Bias and 95% coverage for latent intercept factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values> 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

Figure C 28: Parameter estimation bias (peb) and mean squared error (MSE) for latent intercept factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

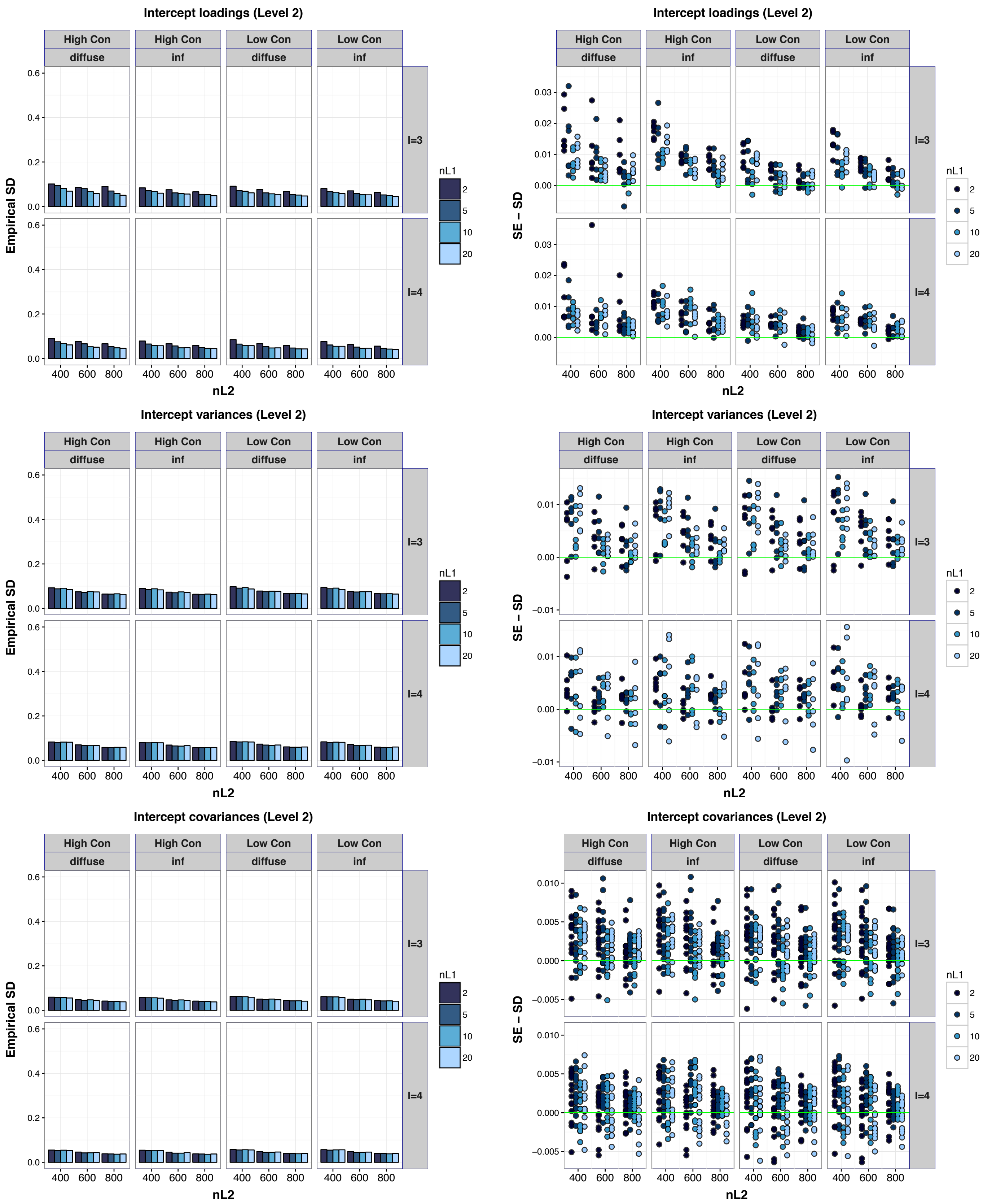

Figure C 29: Empirical SDs and standard error bias (SE - SD) for latent intercept factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

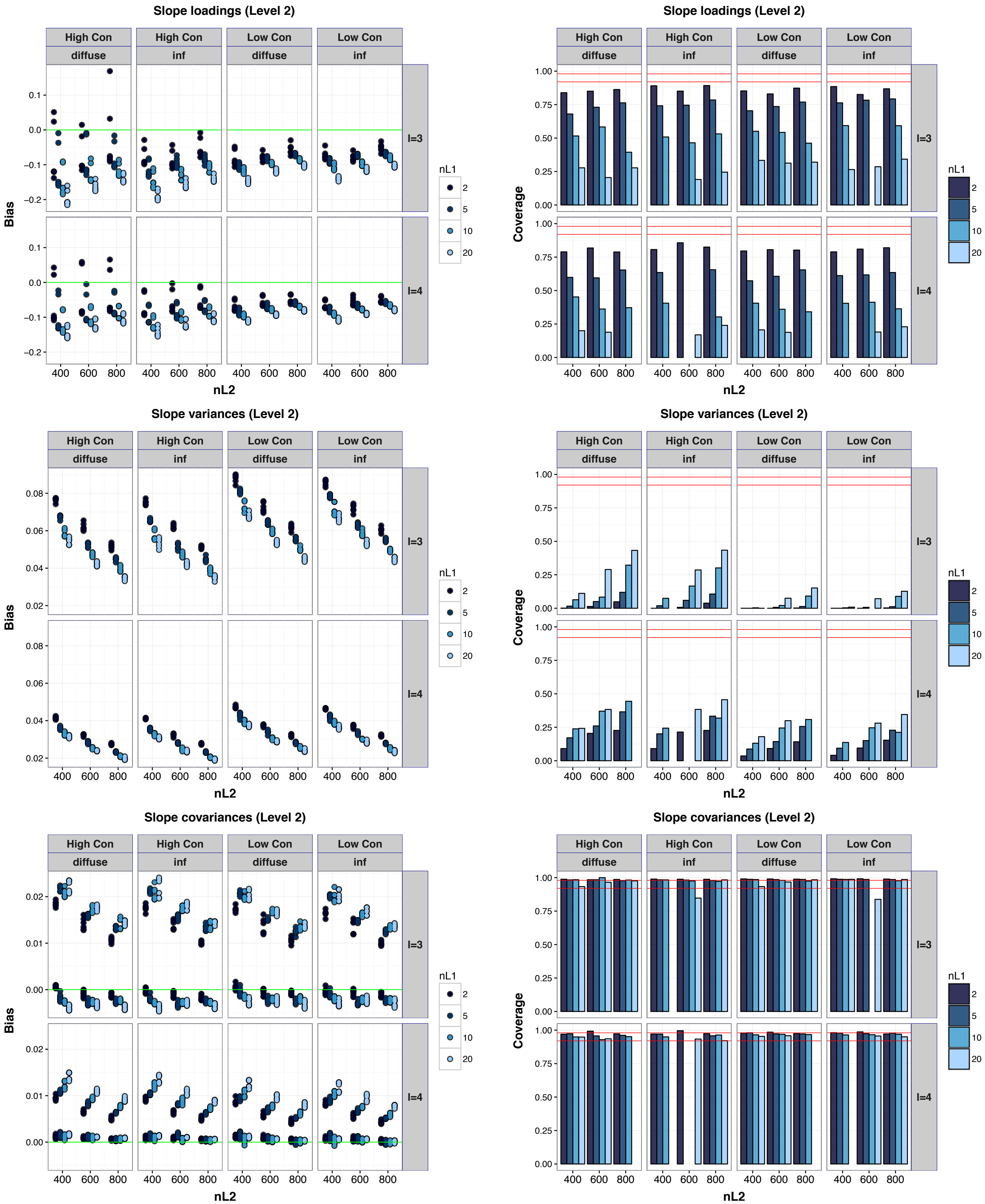

Figure C 30: Bias and 95% coverage for latent slope factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values> 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

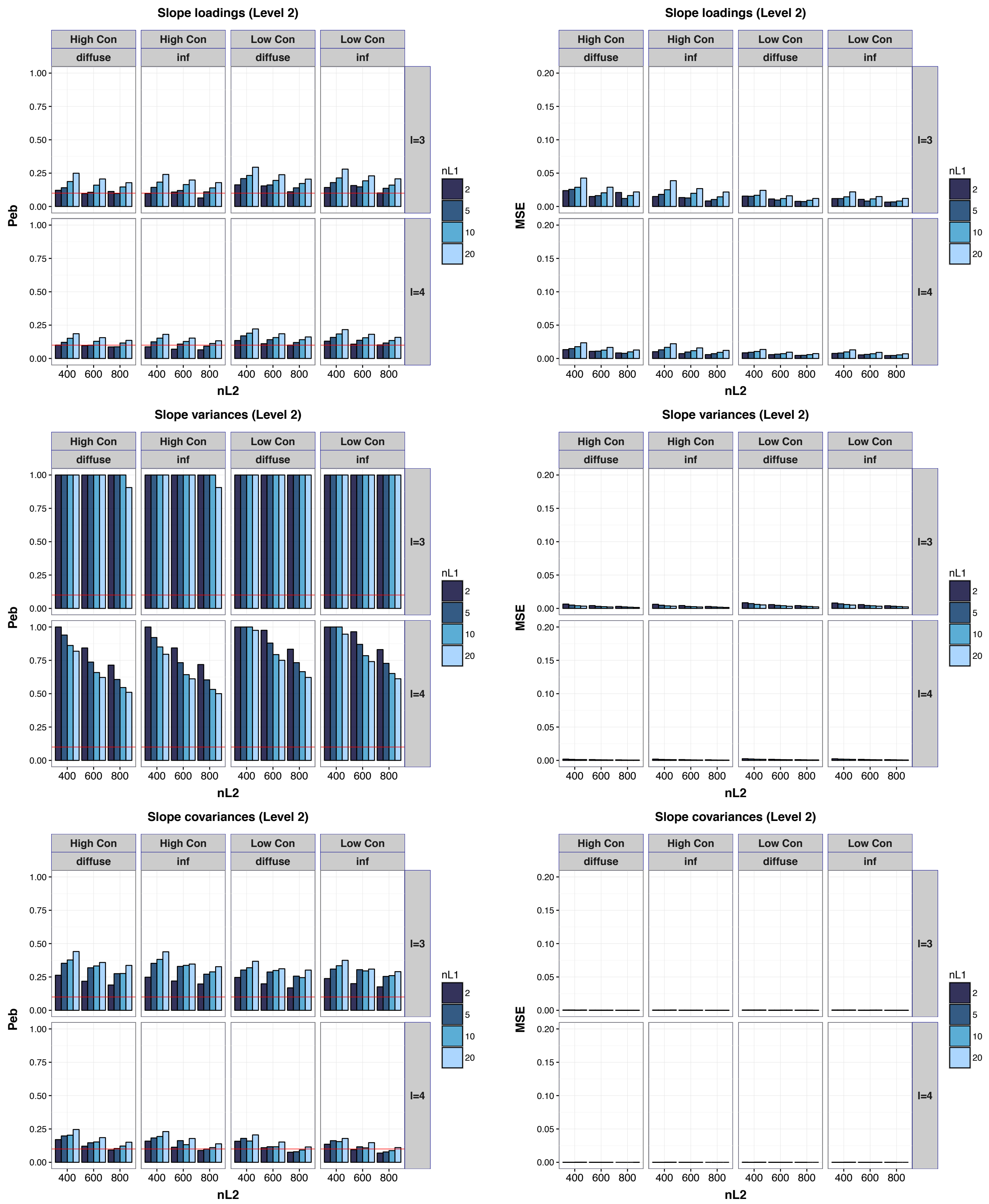

Figure C 31: Parameter estimation bias (peb) and mean squared error (MSE) for latent slope factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

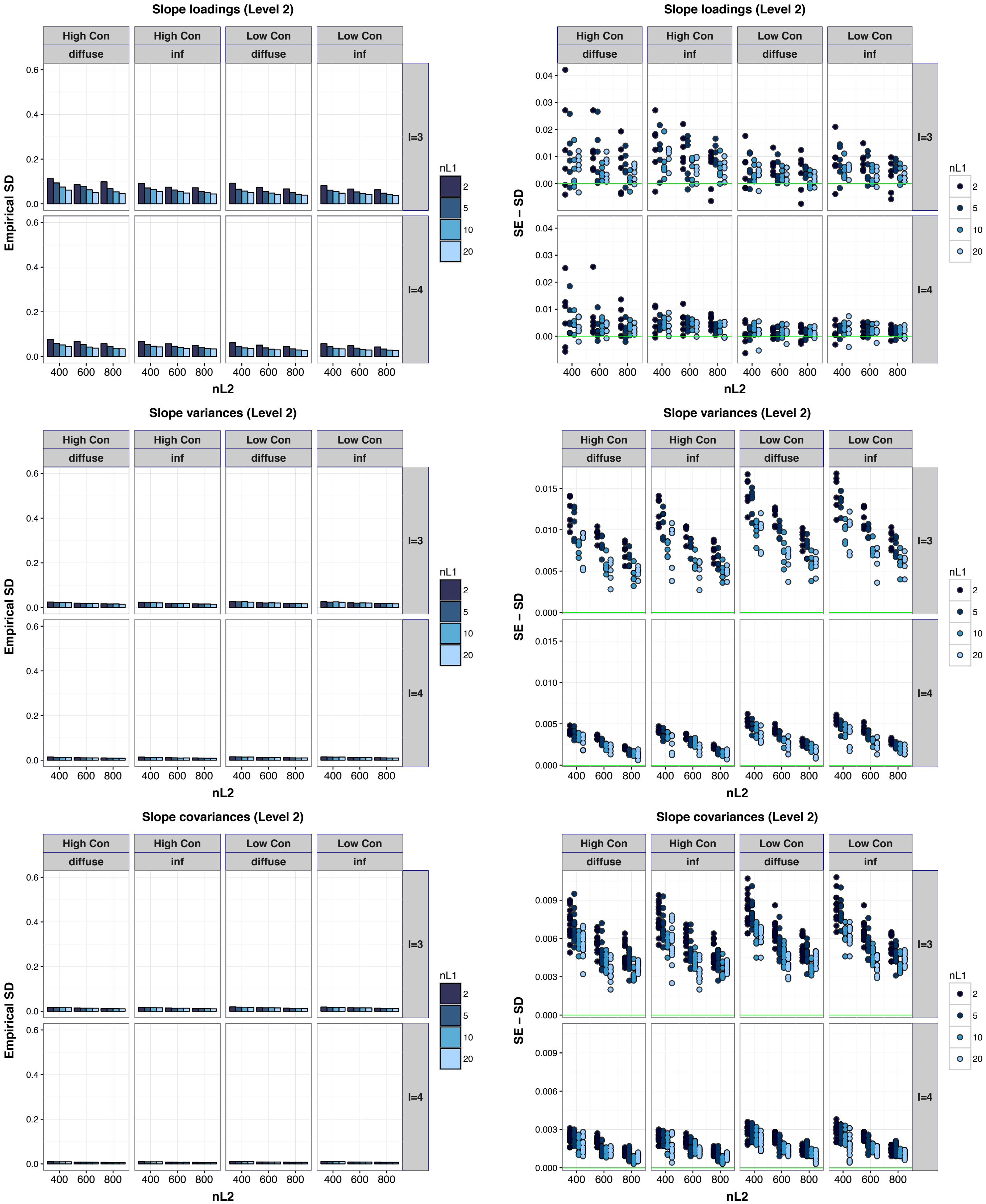

Figure C 32: Empirical SDs and standard error bias (SE - SD) for latent slope factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

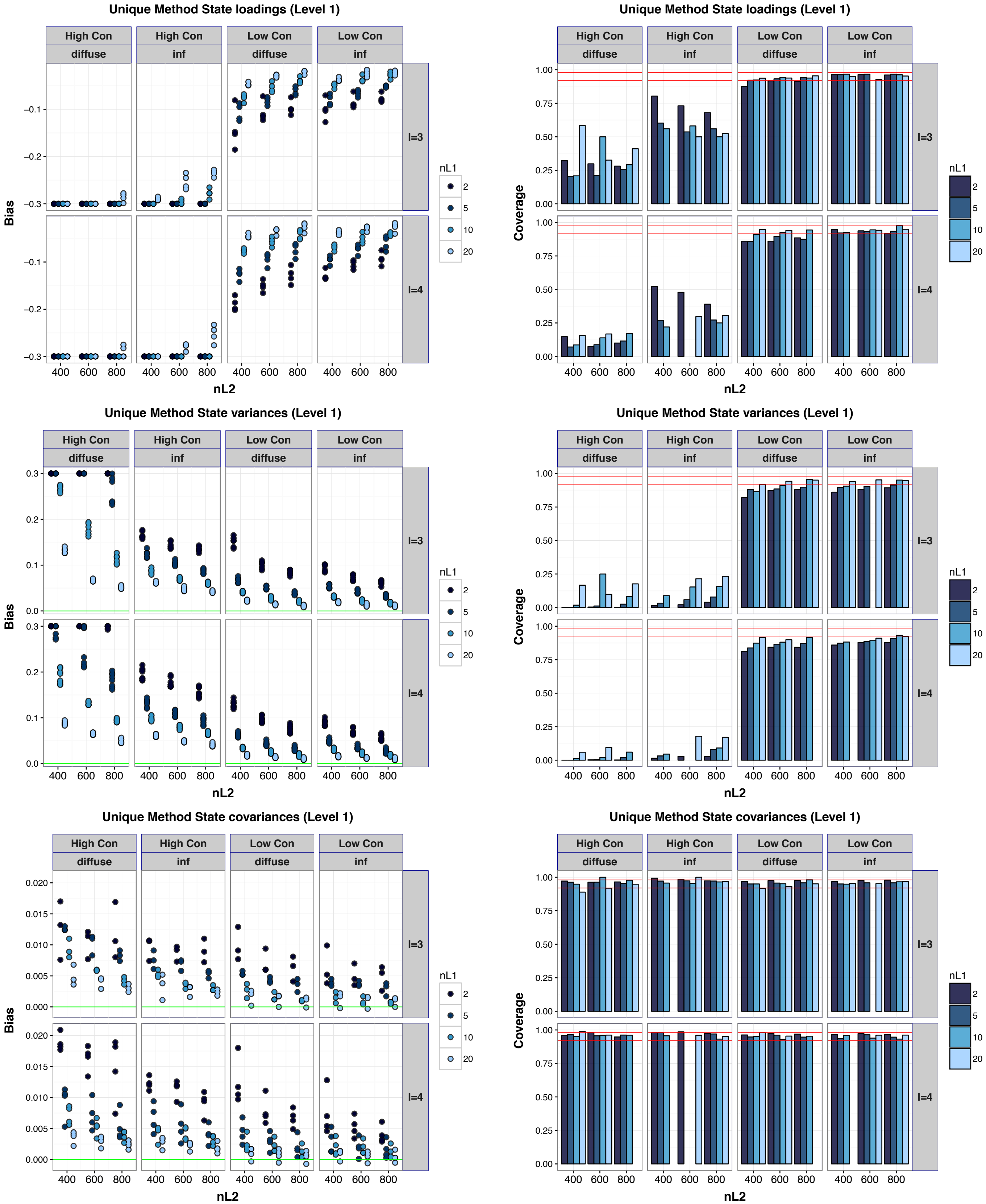

Figure C 33: Bias and 95% coverage for unique method state residual factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

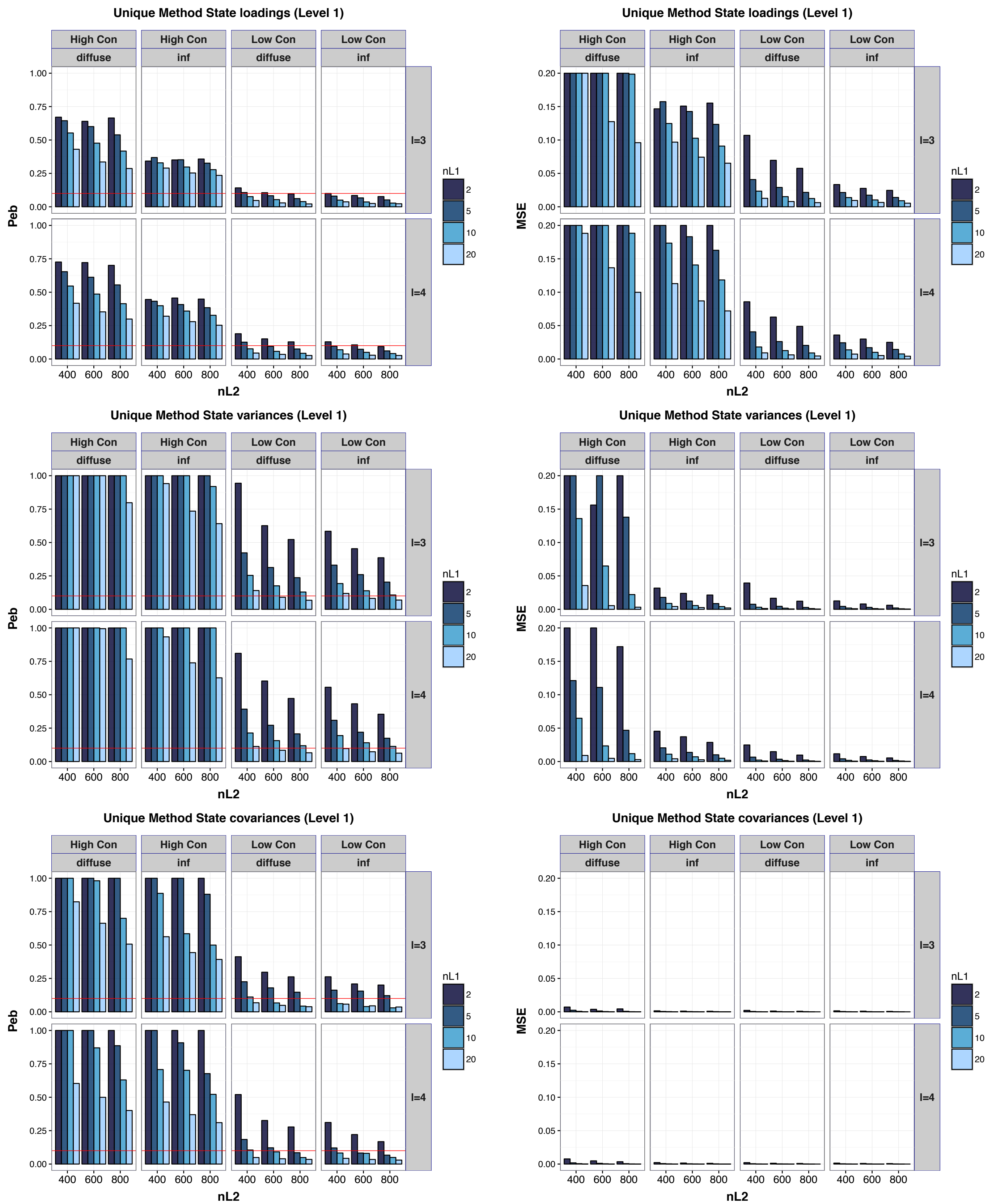

Figure C 34: Parameter estimation bias (peb) and mean squared error (MSE) for unique method state residual factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

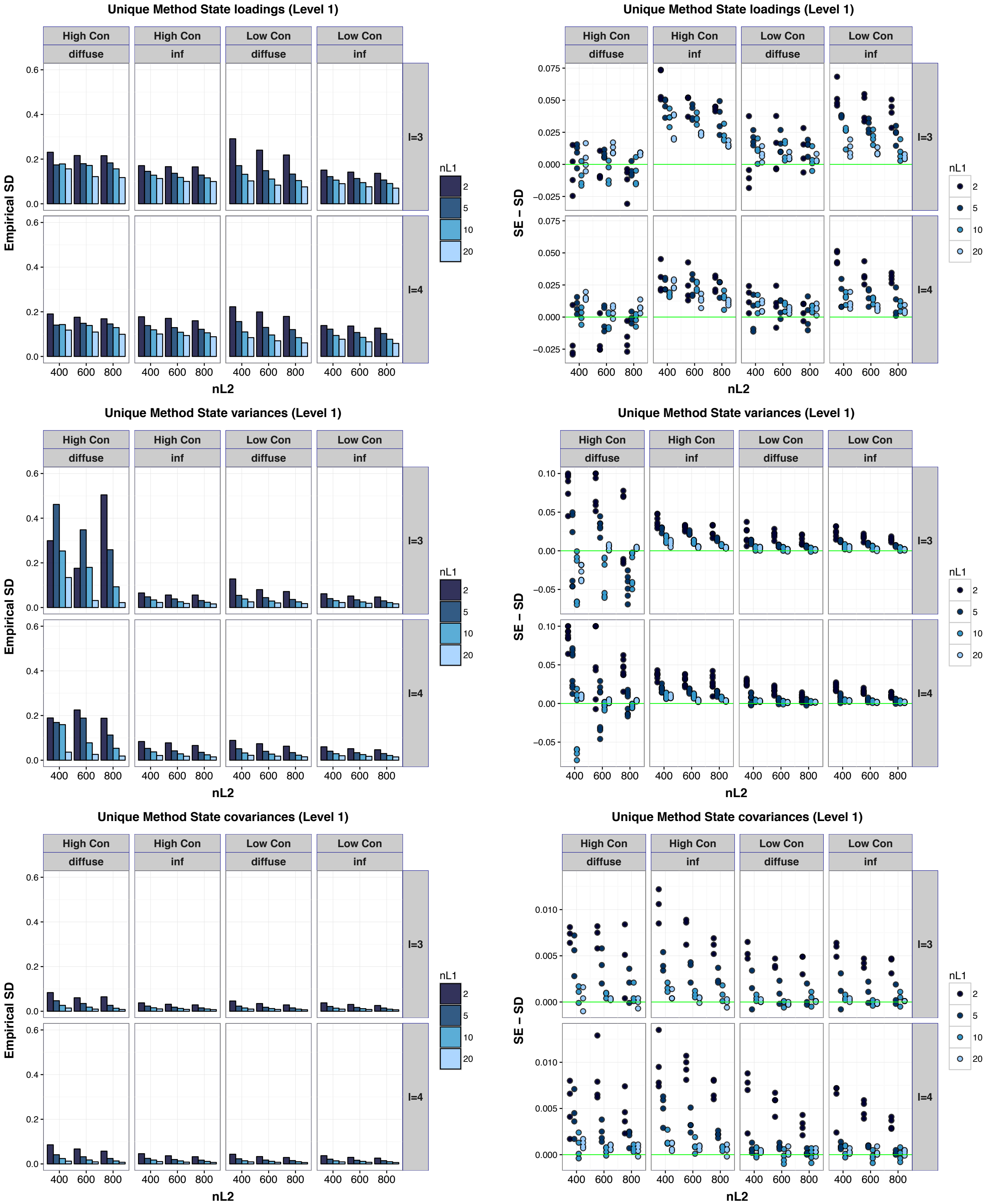

Figure C 35: Empirical SDs and standard error bias (SE - SD) for unique method state residual factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

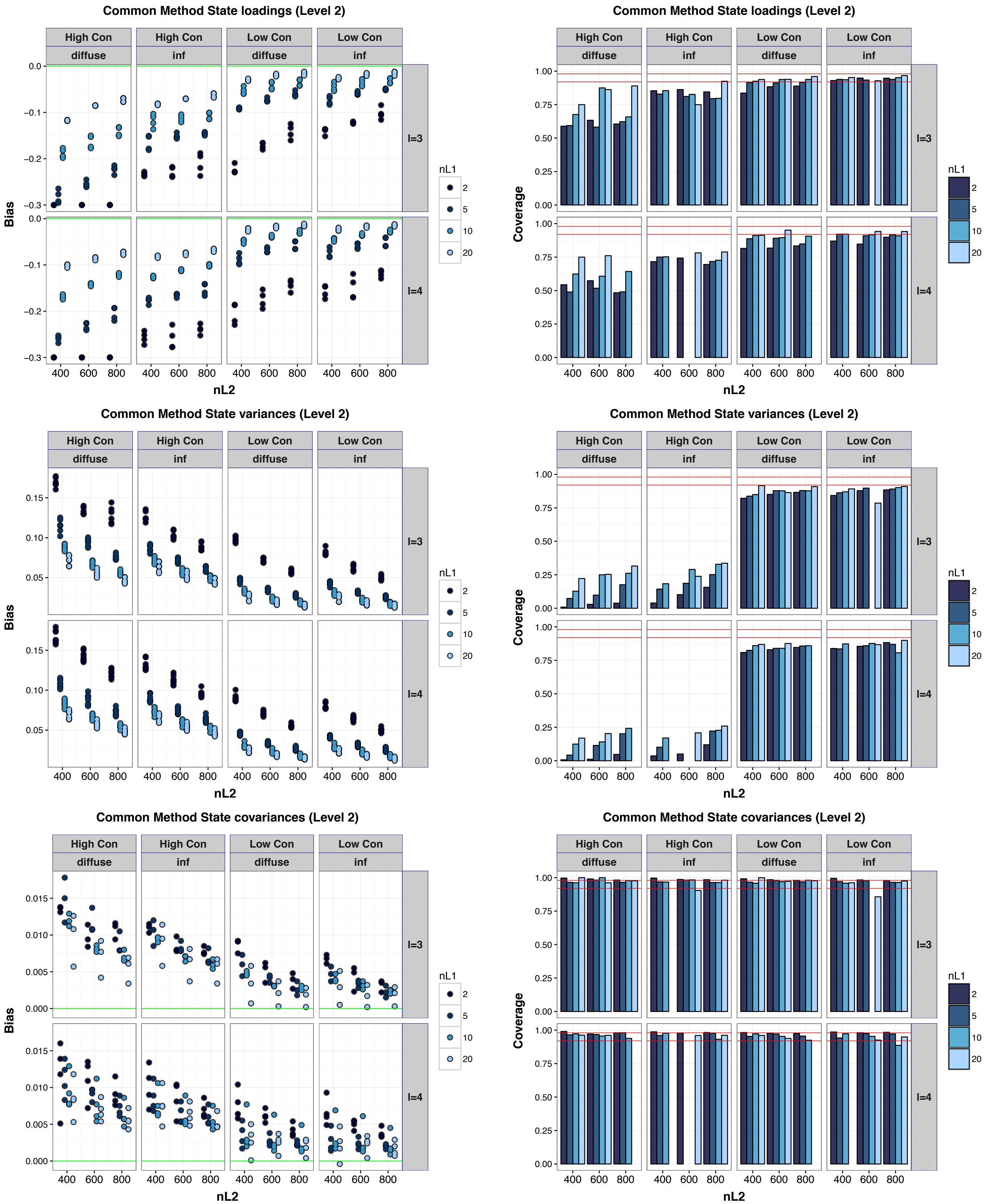

Figure C 36: Bias and 95% coverage for common method state residual factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

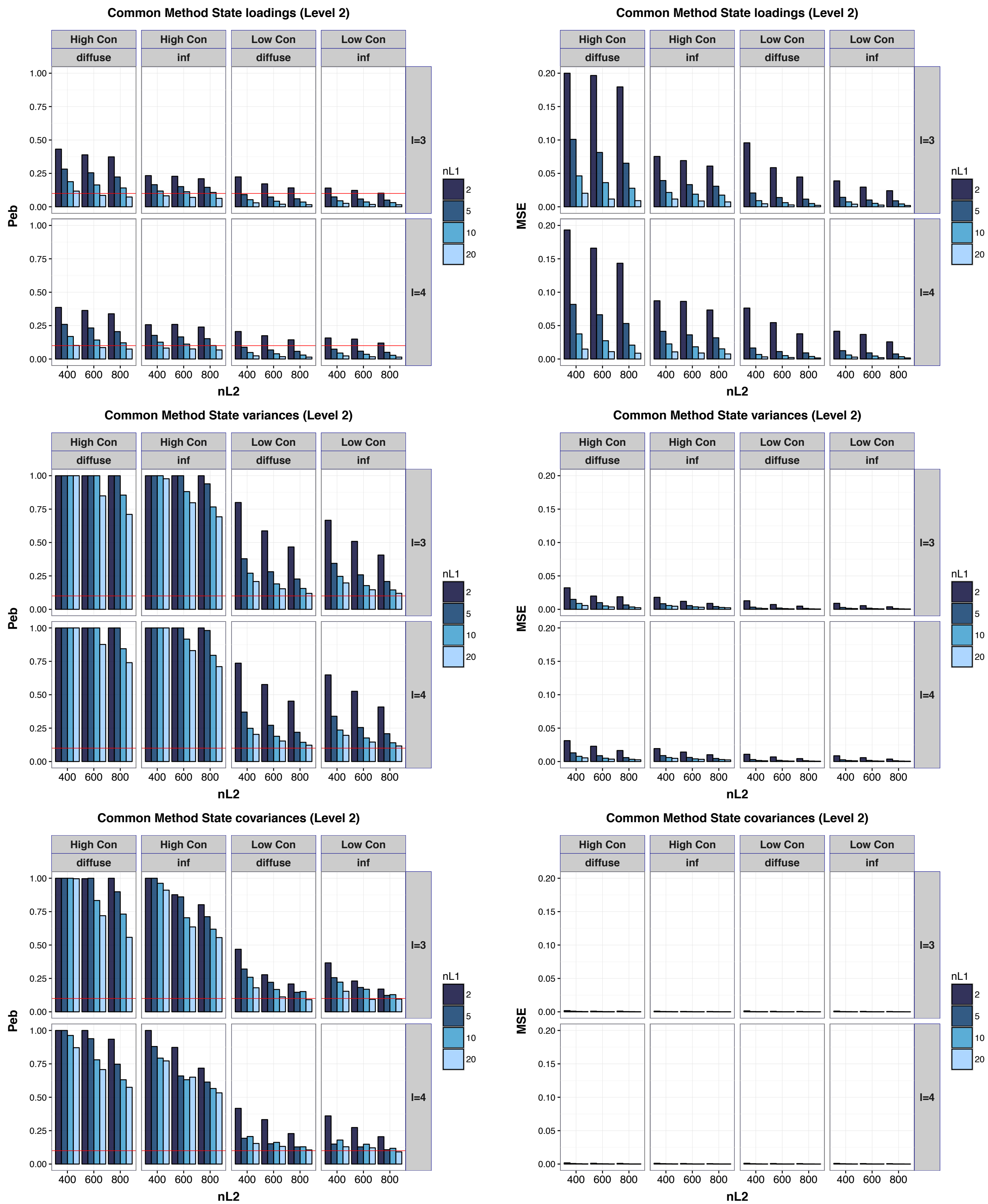

Figure C 37: Parameter estimation bias (peb) and mean squared error (MSE) for common method state residual factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

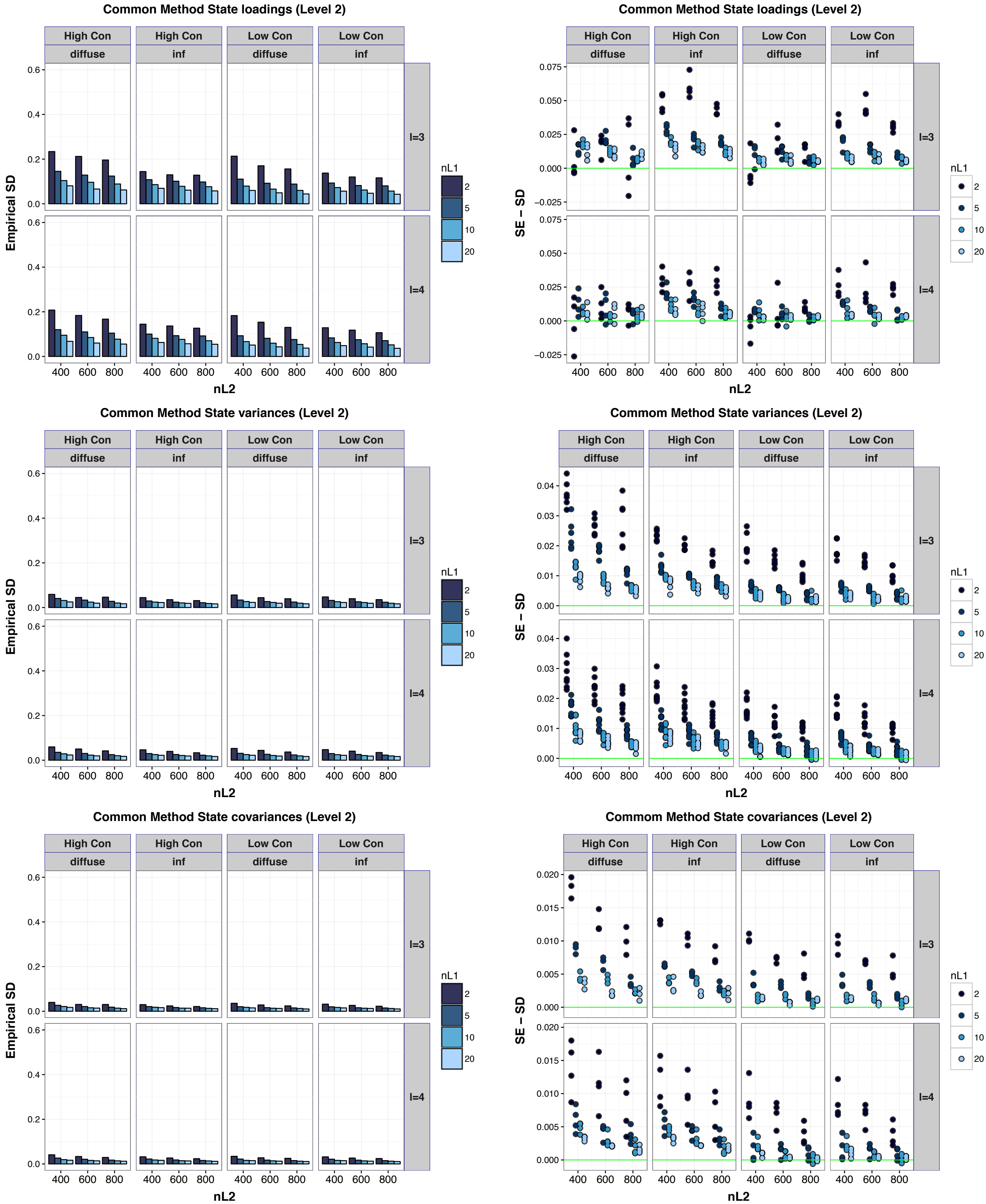

Figure C 38: Empirical SDs and standard error bias (SE - SD) for common method state residual factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

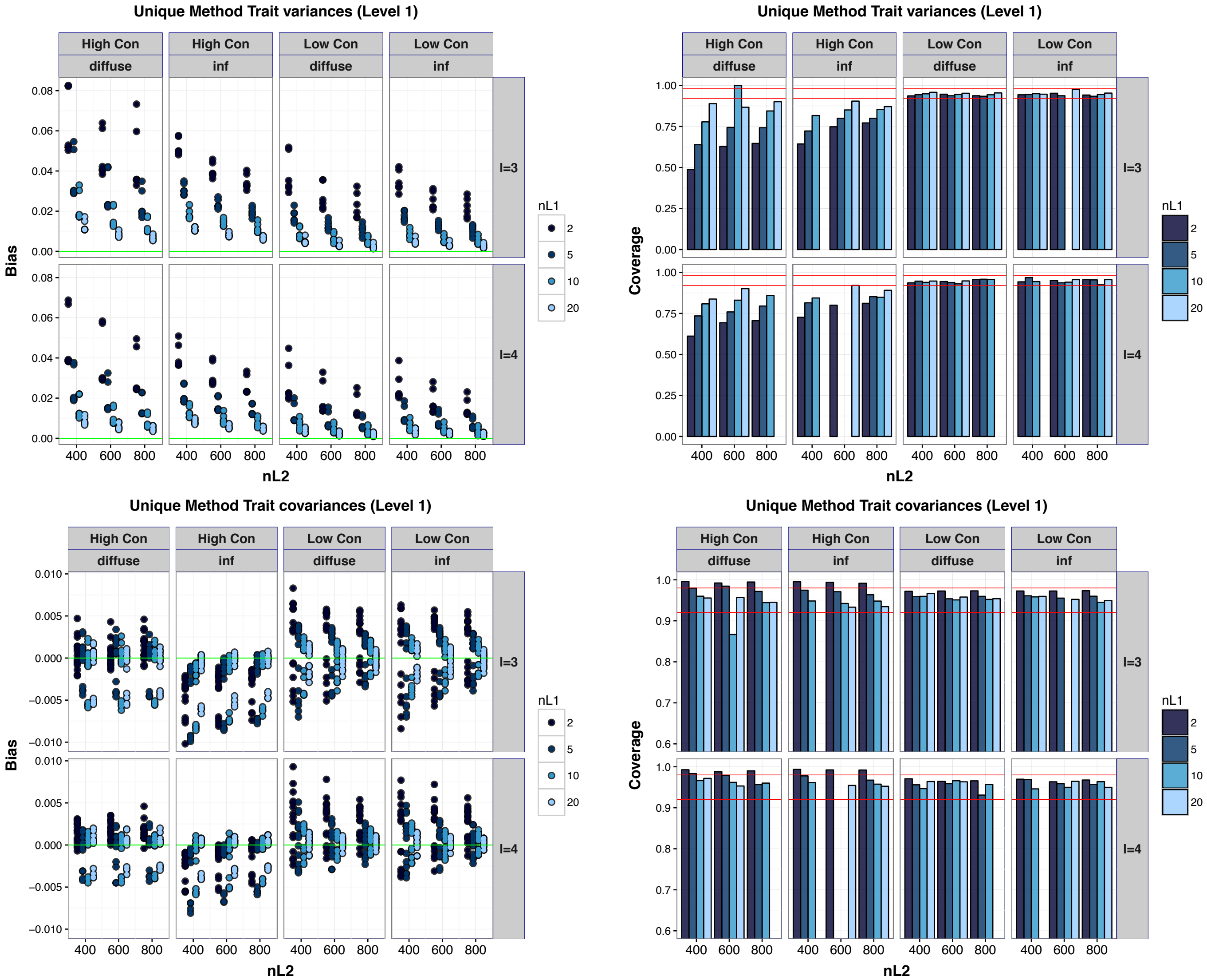

Figure C 39: Bias and 95% coverage for unique method trait factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

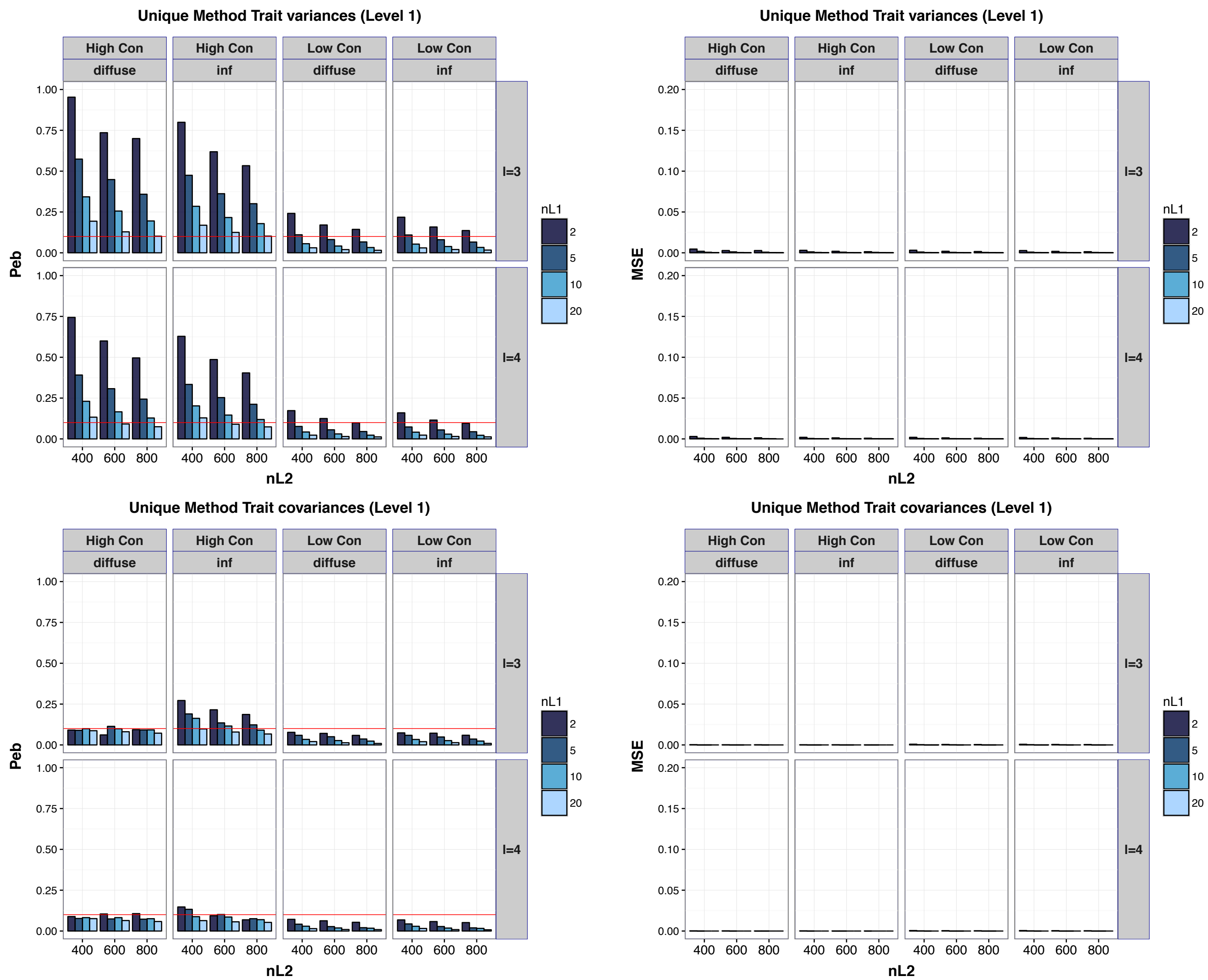

Figure C 40: Parameter estimation bias (peb) and mean squared error (MSE) for unique method trait factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

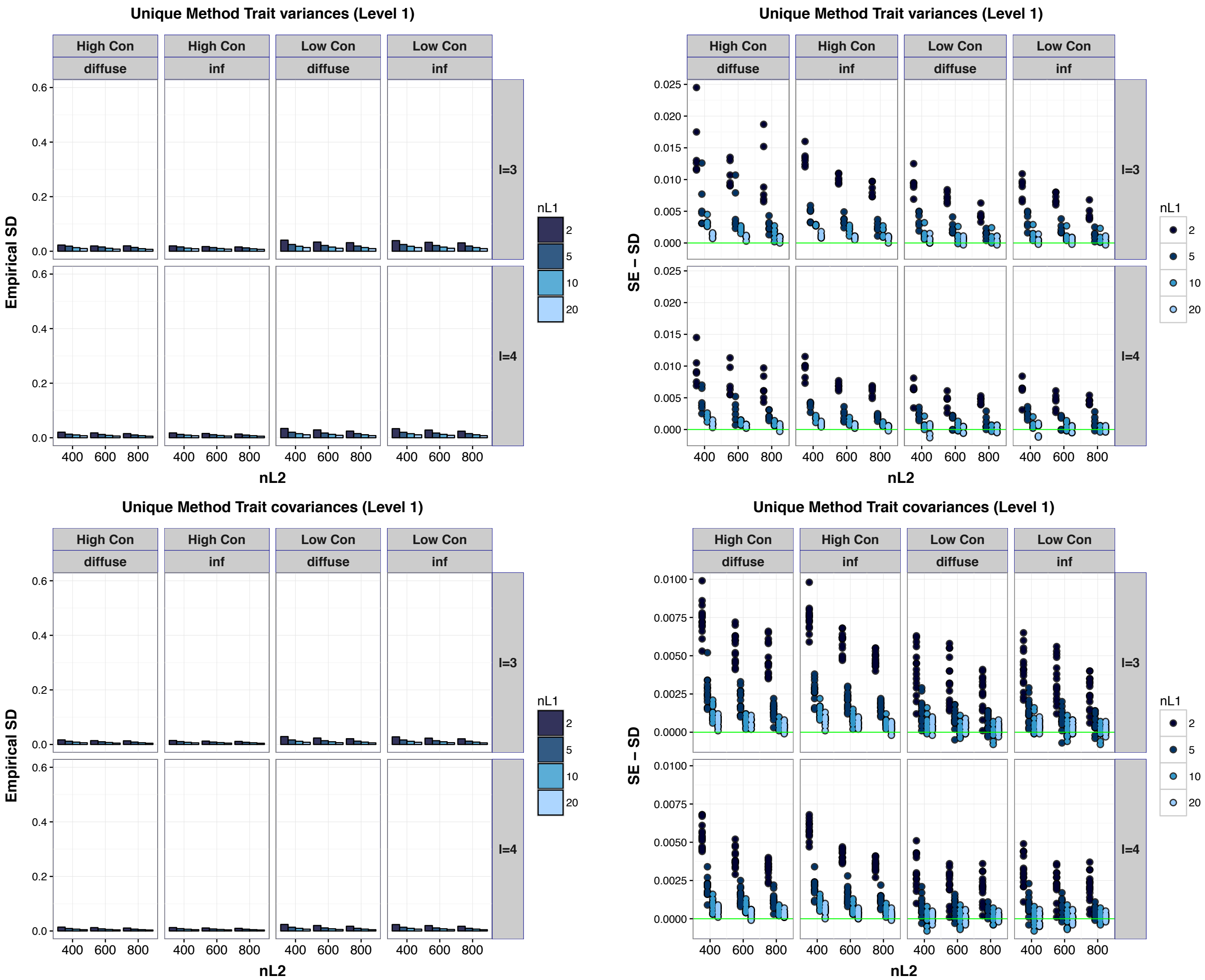

Figure C 41: Empirical SDs and standard error bias (SE - SD) for unique method trait factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences $>$ 0.1 or $<$ -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

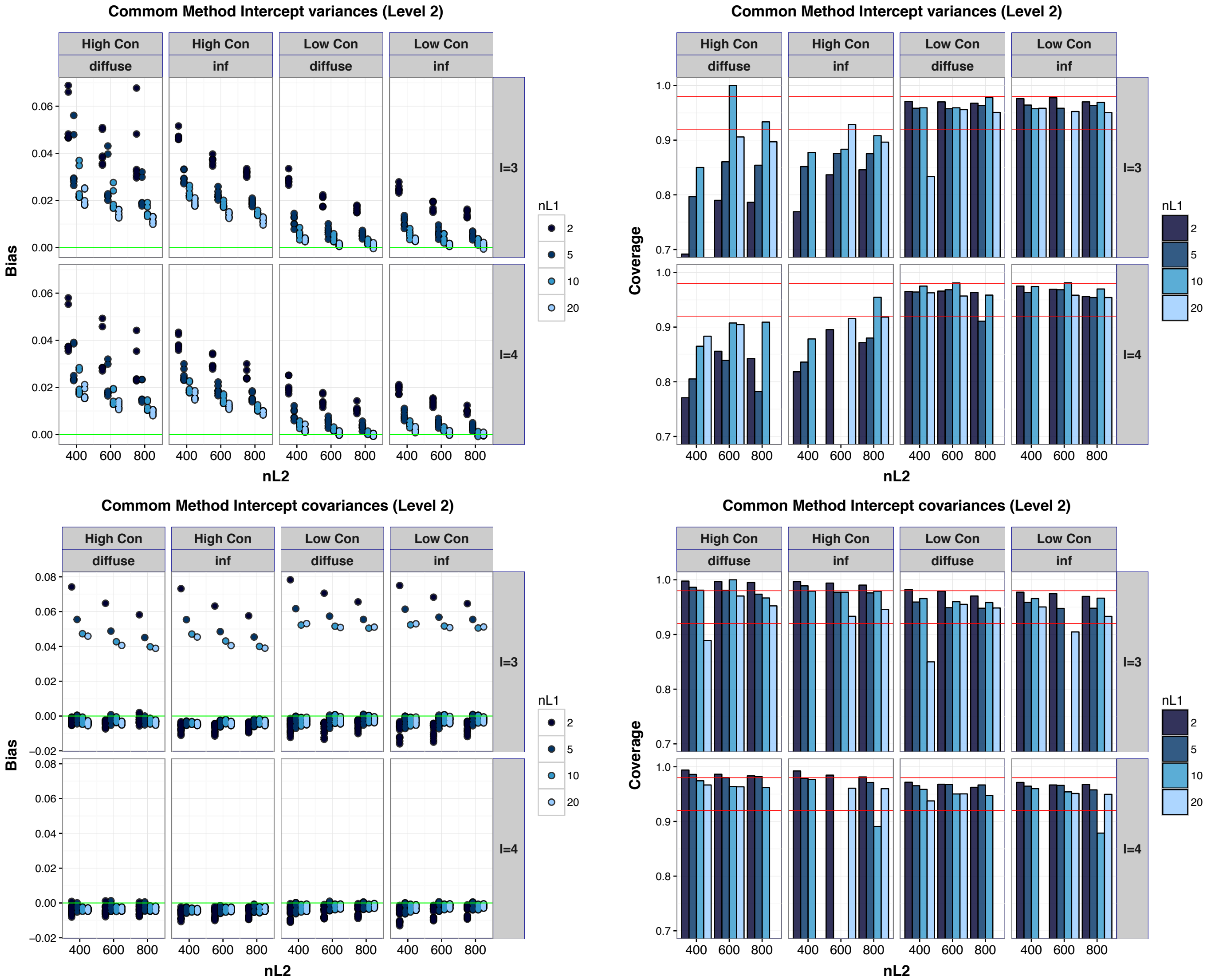

Figure C 42: Bias and 95% coverage for common method intercept factors in the LGC-Com GRM with two constructs. In the plots in the left column, each point represents the bias of a single parameter in the respective condition, as calculated by Equation (7.2.1). Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, to enhance readability of the plot. Note that these values solely occurred in conditions suffering from high rates of non-convergence. Coverage values were averaged over parameters belonging to one parameter class. Coverage values could be calculated based on a reduced number of replications only. These numbers correspond to the number of available PSR values per condition and are displayed in Figure 7.9 (red numbers in the figure). Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Coverage should fall within the region between the two red lines. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

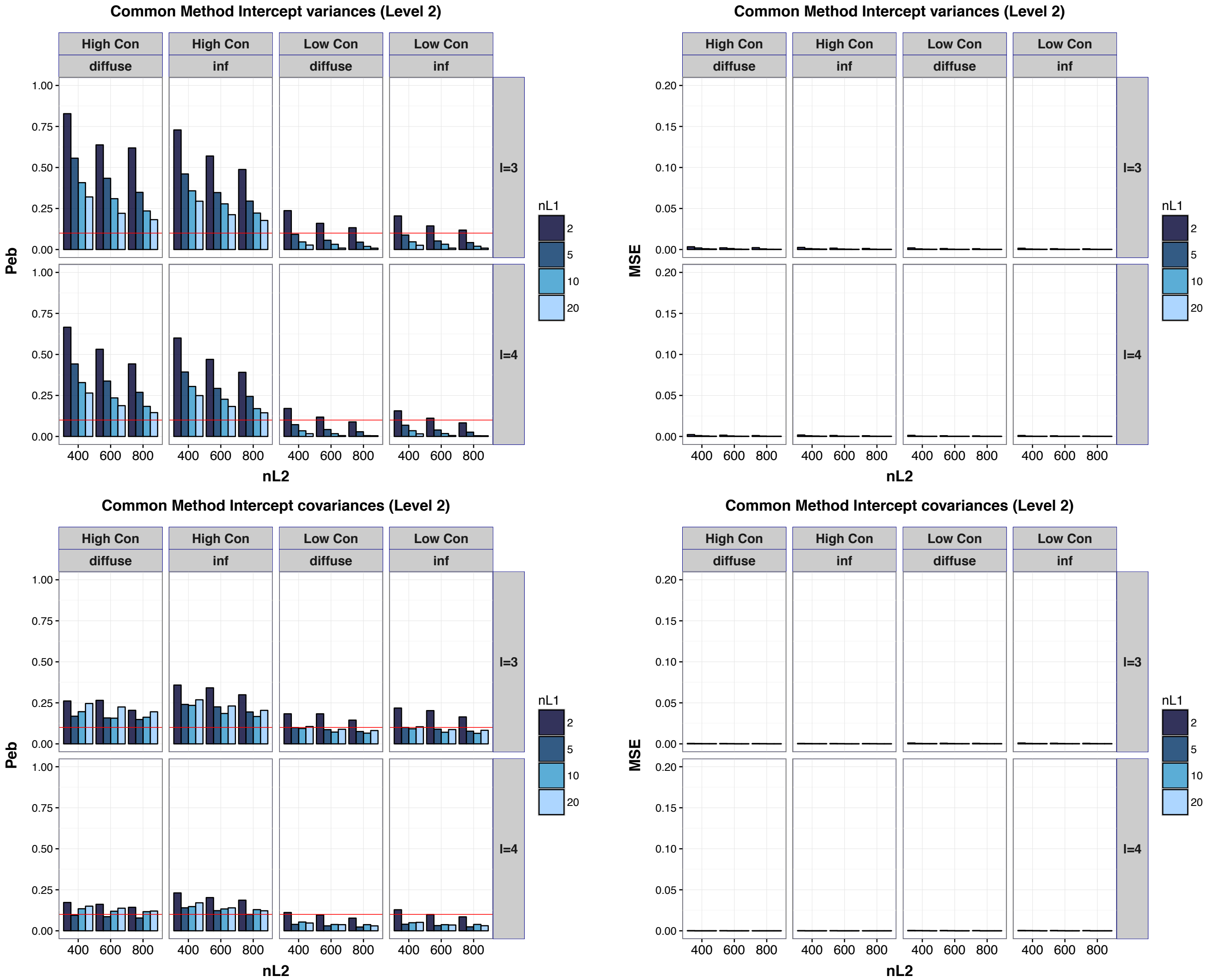

Figure C 43: Parameter estimation bias (peb) and mean squared error (MSE) for common method intercept factors in the LGC-Com GRM with two constructs. Peb, as calculated by Equation (7.2.2), and MSE values are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2 to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value of 0.1 for acceptable peb-values. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

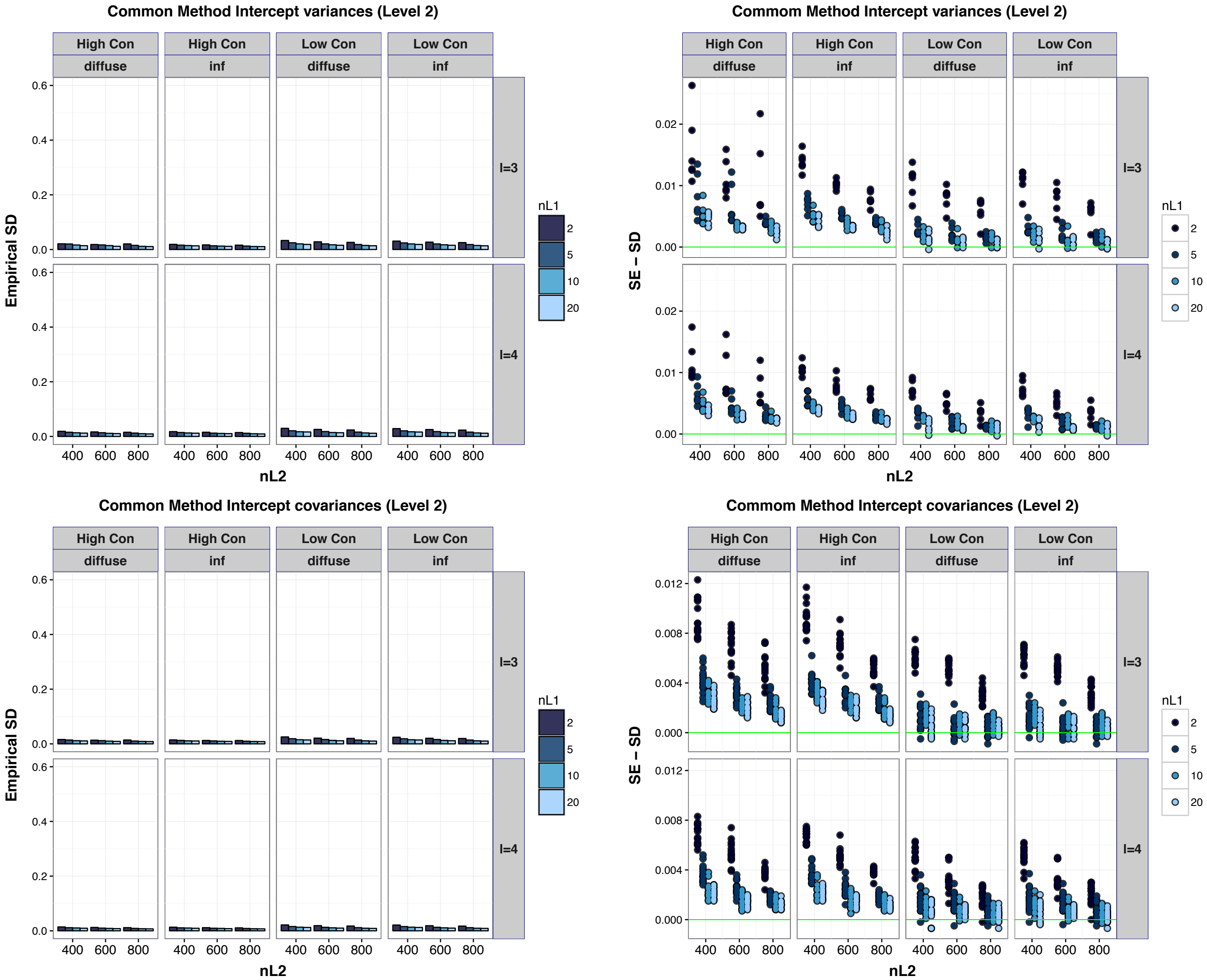

Figure C 44: Empirical SDs and standard error bias (SE - SD) for common method intercept factors in the LGC-Com GRM with two constructs. In the plots in the right column, each point represents the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equation (7.2.3). Empirical SDs were averaged over parameters of one parameter class. Empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

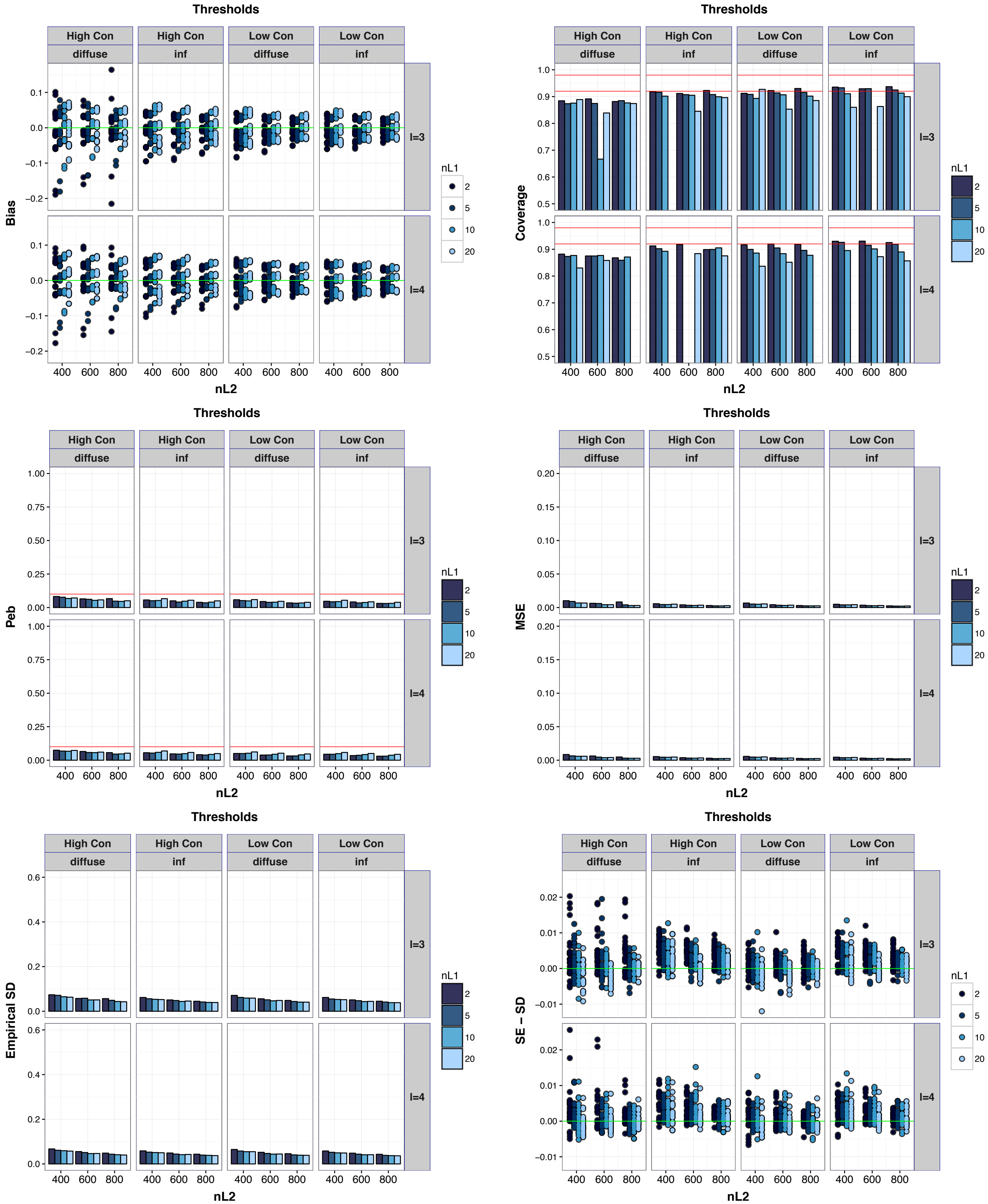

Figure C 45: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the threshold parameters in the LGC-Com GRM with two constructs. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Peb values > 1 were set to 1 and MSE values > 0.2 were set to 0.2, Bias values > 0.3 and < -0.3 were set to 0.3 and -0.3, respectively, empirical SDs > 0.6 were set to 0.6, and SE - SD differences > 0.1 or < -0.1 were set to 0.1 and -0.1, respectively, to enhance readability of the plot. Note that this only concerned conditions suffering from high rates of non-convergence. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.

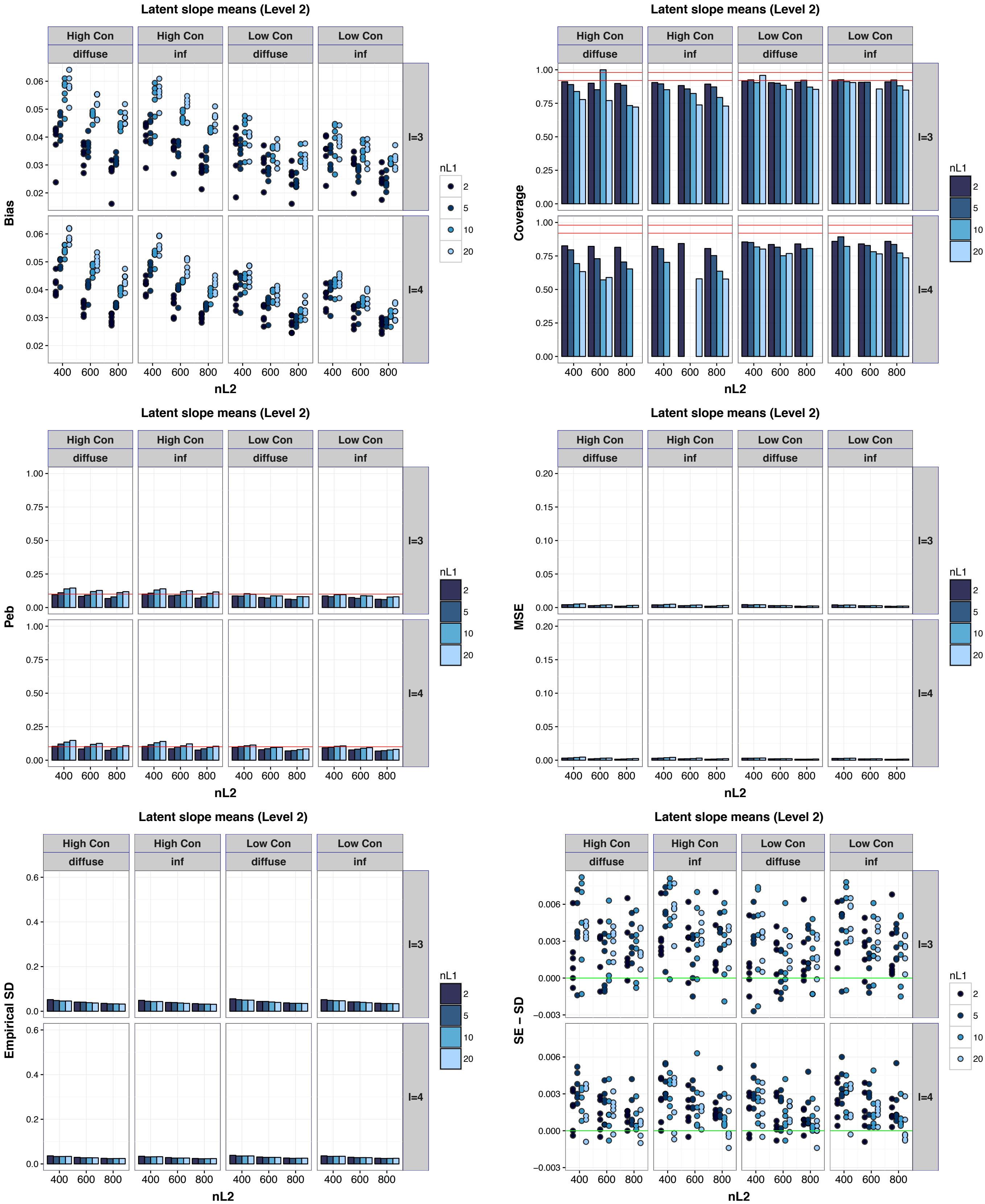

Figure C 46: Bias, 95% coverage, parameter estimation bias (peb), mean squared error (MSE), empirical SDs and standard error bias (SE - SD) for the latent slope means in the LGC-Com GRM with two constructs. Points in the dot plots represent the bias or the difference between empirical SDs and posterior SDs of a single parameter in the respective condition, as calculated by Equations (7.2.1) and (7.2.3). Peb, as calculated by Equation (7.2.2), MSE values and empirical SDs are averaged values over parameters belonging to one parameter class. Note that the limits of the y-axes may differ across plots. The red line indicates the cut-off value for acceptable coverage or peb values. Green lines at zero indicate the point of no bias. Diffuse: diffuse prior condition; High Con: high consistency condition; inf: informative prior condition; l: number of measurement occasions; Low Con: low consistency condition; nL1: number of level-1 observations; nL2: number of level-2 observations.