Model Selection Methods for Panel Vector Autoregressive Models

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft
doctor rerum politicarum
(Dr. rer. pol.)

am Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Annika Schnücker

Berlin, 2018

Erstgutachter: Prof. Dr. Helmut Lütkepohl
Freie Universität Berlin und DIW Berlin

Zweitgutachter: Prof. Dr. Dieter Nautz
Professur für Ökonometrie
Freie Universität Berlin

Tag der Disputation: 11. Juli 2018

In memory of Hedwig

Acknowledgments

First and foremost, I am grateful for the support, advice, and encouragement of my supervisor Helmut Lütkepohl. He gave me the opportunity to work independently on my preferred topics and provided constant guidance. I would also like to thank my second supervisor, Dieter Nautz, for his support and comments. I am very thankful to Tomasz Woźniak for giving me the opportunity to visit the University of Melbourne, for being an extremely attentive and kind host, and for fruitful discussions. I thank my co-author Gregor von Schweinitz for the productive and inspiring cooperation. I am also grateful for the constructive questions and comments of the participants of the regular Seminar Empirical Macroeconomics of the Freie Universität Berlin.

I wrote this dissertation as a member of the DIW Berlin Graduate Center. I am grateful for the support of the GC and the great environment. In particular, I thank the GC macro PhD students for hours of discussions and my GC 2013 cohort for a memorable first year. I thank the GC team and especially Juliane Metzner and Georg Weizsäcker for their support and understanding particularly during the job market.

I am grateful to my officemates and friends, Niels Aka and Hedwig Plamper, for enjoyable conversations beyond economics and GC work. I also thank Pablo Anaya Longaric, Daniel Bierbaumer, Caterina Forti Grazzini, Marie Le Mouel, and Ulrich Schneider for uncountable lunches, coffee breaks, and their friendship outside the DIW.

I thank my family for their encouragement, their belief in me, and for fully endorsing my decisions and priorities. Above all, I am extremely grateful for the endless support and patience of Georg.

Berlin, May 2018 *Annika Schnücker*

Erklärung zu Ko-Autorenschaften

Diese Dissertation besteht aus drei (Arbeits-)Papieren, von denen eines in Zusammenarbeit mit einem Koautor entstanden ist. Der Eigenanteil an Konzeption, Durchführung und Berichtsabfassung der Kapitel lässt sich folgendermaßen zusammenfassen:

- Annika Schnücker:

  "Restrictions Search for Panel Vector Autoregressive Models"

  Eigenanteil: 100 Prozent

- Annika Schnücker:

  "Penalized Estimation of Panel Vector Autoregressive Models: A Lasso Approach"

  Eigenanteil: 100 Prozent

- Annika Schnücker und Gregor von Schweinitz:

  "International Monetary Policy Transmission"

  Eigenanteil: 50 Prozent

Liste der Vorpublikationen

Working Papers

Schnücker, A., 2016. Restrictions Search for Panel VARs. DIW Discussion Papers 1612, Berlin
Vorpublikation von Kapitel 1

Contents

List of Figures

List of Tables

List of Abbreviations

APD	absolute percentage deviation
BMA	Bayesian model averaging
BoE	Bank of England
BVAR	Bayesian vector autoregressive
CA	Canada
CC	cross-sectional shrinkage approach
CPI	consumer price index
CSH	cross-sectional heterogeneities
DE	Germany
DGP	data generating process
DI	dynamic interdependencies
DK	Denmark
EA	euro area
ECB	European Central Bank
EBP	excess bond premium
EME	emerging markets economies
ES	Spain
FAVAR	factor augmented vector autoregressive
Fed	Federal Reserve System
FOMC	Federal Open Market Committee
FR	France
GDP	gross domestic product
glasso	graphical least absolute shrinkage and selection operator
GLS	generalized least squares
GR	Greece
GVAR	global vector autoregressive
IE	Ireland
IMF	International Monetary Fund
IP	industrial production

IR	interest rate
IT	Italy
JP	Japan
lasso	least absolute shrinkage and selection operator
lassoPVAR	least absolute shrinkage and selection operator for panel vector autoregressive models
LS	least squares
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MSE	mean squared error
MSFE	mean squared forecast error
OECD	Organisation for Economic Co-Operation and Development
OLS	ordinary least squares
PL	predictive density
PT	Portugal
PVAR	panel vector autoregressive
REER	real effective exchange rate
SA	seasonally adjusted
SI	static interdependencies
S^4	stochastic search specification selection
SSSS	stochastic search specification selection
SSVS	stochastic search variable selection
SSVSP	stochastic search variable selection for panel vector autoregressive models
SUR	seemingly unrelated regression
SVAR	structural vector autoregressive
UK	United Kingdom
UN	unemployment rate
US	United States
VAR	vector autoregressive

Summary

Multi-country dynamic time series models, called panel vector autoregressive (PVAR) models, allow for multilateral cross-border linkages and country-specific dependencies among variables. Thus, these models are excellent tools for macroeconomic spillover analyses. However, as they jointly model multiple variables of several countries, the dimensionality of unrestricted PVAR models is large and the estimation feasibility is thus not guaranteed with standard methods. Hence, model selection techniques which restrict PVAR models and thereby reduce the dimensionality of the models are necessary to ensure the estimation feasibility. Chapters 1 and 2 of this thesis propose Bayesian and classical selection methods for PVAR models which search for restrictions supported by the data and which take specific panel properties into account. Furthermore, theoretical arguments for commonly used recursive structural identification in multi-country models are often insufficient. The third chapter analyzes international monetary policy spillovers in a three-country vector autoregressive model using external instruments to identify monetary policy shocks.

The first chapter introduces a Bayesian selection prior for PVAR models. The proposed selection prior allows for a data-based restrictions search ensuring the estimation feasibility. The prior is specified as a mixture distribution which allows to shrink parameters to restrictions or to estimate them freely. The prior specification differentiates between domestic and foreign variables by searching for zero restrictions on lagged foreign variables and for homogeneity across countries for coefficients of domestic variables. The prior, thereby, allows for a flexible panel structure and a restrictions search on single elements. Furthermore, the prior searches for restrictions on the covariance matrix. A Monte Carlo simulation shows that the selection prior outperforms alternative estimators for flexible panel structures in terms of mean squared errors measuring the deviation of the parameter estimates from the true values. Furthermore, a forecast exercise for G7 countries demonstrates that forecast performance improves for the proposed prior focusing on sparsity in form of no dynamic interdependencies.

The second chapter proposes a new lasso (least absolute shrinkage and selection operator) for estimating PVAR models. The penalized regression ensures the feasibility of the estimation by specifying a shrinkage penalty that accommodates time series and cross section characteristics. It thereby accounts for the inherent panel structure within the data. Furthermore, using the weighted sum of squared residuals as the loss function enables the lasso for PVAR models to take into account correlations between cross-sectional units in the penalized regression. The specification of the penalty term allows to establish the asymptotic oracle properties. Given large and sparse models, simulation results point towards advantages of using the lasso for PVAR models over ordinary least squares estimation, standard lasso techniques as well as Bayesian estimators in terms of mean squared errors measuring the deviation of the estimates from their true values and forecast accuracy. Empirical forecasting applications with up to ten countries and four variables support these findings.

The third chapter assesses the international macroeconomic spillover effects of monetary policy shocks for the United States, the United Kingdom, and the euro area. The Bayesian proxy three-country structural vector autoregressive model accounts for international interdependencies and traces the dynamic cross-border responses of macroeconomic variables to monetary policy shocks identified with external instruments. The instruments for monetary policy surprises capture changes in high frequency government bond future contracts around policy announcement dates. The results provide no evidence for cross-border macroeconomic effects.

Keywords: Model selection, multivariate time series, large vector autoregressions, multi-country models, panel data, selection prior, lasso, penalized regression, shrinkage estimation, international monetary policy spillover, external instrument identification, high-frequency identification

JEL Classification: C11, C32, C33, C52, E52, E58, F42

Zusammenfassung

Dynamische Zeitreihenmodelle für mehrere Länder, genannt Panel vektorautoregressive (PVAR) Modelle, können gleichzeitig multilaterale internationale Abhängigkeiten und länderspezifische Eigenschaften modellieren. Damit eignen sich diese Modelle hervorragend zur Analyse von globalen, makroökonomischen Entwicklungen und grenzüberschreitenden Effekten. PVAR Modelle integrieren Variablen mehrerer Ländern in ein gemeinsames Modell. Somit ist die Dimensionalität der nicht restringierten Modelle so groß, dass diese häufig nicht mehr mit Standardmethoden geschätzt werden können. Um die Schätzbarkeit der Modelle zu garantieren, ist es notwendig, PVAR Modelle mithilfe Methoden der Modellselektion zu beschränken und damit die Dimensionalität der Modelle zu reduzieren. Kapitel 1 und 2 dieser Dissertation führen bayesianische und frequentistische Selektionsmethoden für PVAR Modelle ein, die datenbasierte Restriktionen suchen und dabei spezifische Eigenschaften von Paneldaten berücksichtigen. Darüber hinaus ist die strukturelle Identifizierung von PVAR Modellen aufgrund unzureichender theoretischer Argumente oftmals problematisch. Das dritte Kapitel analysiert internationale Effekte von geldpolitischen Schocks in einem PVAR Modell. Die strukturelle Identifizierung der geldpolitischen Schocks basiert auf externen Instrumenten.

Das erste Kapitel führt einen bayesianischen *selection prior* für PVAR Modelle ein. Die vorgeschlagene a-priori Verteilung ermöglicht eine datenbasierte Suche von Restriktionen, die die Schätzbarkeit des Modells garantieren. Die a-priori Verteilung ist als Mischverteilung spezifiziert, die Parameter gegen Restriktionen schrumpft oder frei schätzt. Die Spezifizierung der a-priori Verteilung unterscheidet zwischen inländischen und ausländischen Variablen, indem nach Null-Restriktionen für ausländische Variablen und Homogenitäten zwischen Ländern für inländische Variablen gesucht wird. Die a-priori Verteilung nimmt somit eine flexible Panelstruktur an und führt eine Restriktionssuche basierend auf einzelnen Variablen durch. Die Ergebnisse von Monte Carlo Simulationen zeigen, dass bei flexibleren Panelstrukturen die mittleren quadratischen Abweichungen der geschätzten Werte vom wahren Wert mit dem vorgeschlagenen *selection prior* geringer sind als bei alternativen Schätzmethoden. Ebenso demonstriert

eine Prognoseanwendung für G7 Länder, dass die Prognosefähigkeit der eingeführten a-priori Verteilung verbessert wird, wenn nach dem Fehlen von dynamischen Abhängigkeiten gesucht wird.

Das zweite Kapitel führt einen neuen *lasso* (*least absolute shrinkage and selection operator*) zur Schätzung von Panel vektorautoregressiven Modellen ein. Dieser regularisierte Regressionsschätzer gewährleistet die Schätzung, indem eine Beschränkung spezifiziert wird, die sowohl Eigenschaften von Zeitreihen- als auch von Querschnittdaten berücksichtigt. Die in den Daten enthaltene Panelstruktur wird somit erfasst. Außerdem berücksichtigt die Spezifizierung der Verlustfunktion in der regularisierten Schätzung als gewichtete Residuenquadratsumme Korrelationen zwischen den Querschnittseinheiten. Die Spezifizierung der Beschränkung erlaubt es zudem, die asymptotischen *Oracle* Eigenschaften nachzuweisen. Die Monte Carlo Simulationen mit großen und sparsamen Modellen demonstrieren die Vorteile des *lasso* für PVAR Modelle gegenüber dem Kleinste-Quadrate-Schätzer, Standardvarianten des *lasso* und weiteren bayesianischen Schätzmethoden. So minimiert der *lasso* für PVAR Modelle die mittleren quadratischen Abweichungen der geschätzten Werte von deren wahren Werten und verbessert die Prognosegenauigkeit. Eine empirische Anwendung zur Prognose mit bis zu zehn Ländern und vier Variablen unterstützt die Ergebnisse.

Das dritte Kapitel untersucht die internationalen makroökonomischen Effekte von geldpolitischen Schocks für die Vereinigten Staaten, Großbritannien und für den Euroraum. Das verwendete bayesianische Proxy strukturelle vektorautoregressive Modell für die drei Länder kann multilaterale globale Verknüpfungen erfassen und zeichnet die dynamischen grenzüberschreitenden makroökonomischen Auswirkungen von geldpolitischen Schocks nach. Die strukturellen geldpolitischen Schocks werden mit externen Instrumenten identifiziert. Die Instrumente erfassen Veränderungen der hochfrequenten Daten für Futures auf Staatsanleihen an Tagen mit geldpolitischen Ankündigungen. Die Resultate zeigen keine Evidenz für grenzüberschreitende makroökonomische Effekte.

Introduction and Overview

Large dynamic multivariate time series models, called large vector autoregressive (large VAR) models, are an important tool for macroeconomic analyses. Policy makers in central banks or government institutions are concerned with questions involving a large number of time series to capture complex macroeconomic dynamics. Moreover, the number of variables of interest can rapidly grow when the focus is on disaggregated data or when a cross-sectional dimension is incorporated. Furthermore, large VARs become increasingly relevant as large datasets for macroeconomic variables are available and computational capacities exist to estimate these models. Among others, Bernanke et al. (2005), Banbura et al. (2010), and Jarociński and Maćkowiak (2017) show that VAR models including numerous time series perform well for forecasting and structural analysis.

However, standard VAR models are limited by the number of variables which can be included. These models therefore face a trade-off between estimation precision and omitted variable bias. Including many variables in a VAR model with limited number of time series observations reduces the estimation precision. The number of observations can easily be lower than the number of parameters to estimate. In that case, standard estimation techniques such as ordinary least squares are not feasible any more. Yet, excluding variables from the model to reduce the dimensionality can cause an omitted variables bias. This bias can have an impact on conclusions of structural analyses.[1] Additionally, the decision which variables to include in the model is often ad hoc as in many cases economic theory may provide insufficient guidance.

Different model selection techniques for large VAR models overcome the issue of estimation feasibility by reducing the model dimension. Among them, Bayesian selection methods reduce the parameter space by using specific prior distributions which shrink coefficients. Classical shrinkage methods such as ridge regression or the least absolute shrinkage and selection operator constrain ordinary least squares estimation by introducing penalty terms. These penalties lead to a shrinkage of coefficient estimates or set parameters equal to zero. Moreover, the use of latent factors in a large model enables

[1] Kilian and Lütkepohl (2017), Chapter 16, elaborate further on this issue.

to extract certain information of multiple variables in a lower number of factors.

This thesis focuses on model selection methods for specific large VAR models, namely, large dynamic multi-country time series models, called panel vector autoregressive (PVAR) models. The dimensionality of PVAR models is large as they trace the dynamic interactions of variables of multiple countries in one model. A PVAR model jointly includes multiple variables of several cross-sectional units. The unrestricted PVAR model has three key characteristics. First, it allows for dynamic interdependencies by including lagged endogenous variables of all countries in each equation. Second, the model accounts for cross-country heterogeneities since coefficient matrices are country-specific. Third, it captures static interdependencies as an unrestricted covariance matrix allows for correlation between all error terms of all countries. Typical restrictions for PVAR models ensuring the estimation feasibility are allowing interdependencies to exist only between specific country and variable combinations or assuming that coefficients are homogeneous across economies. This thesis focuses on medium-sized models as the number of countries included does not exceed ten. These medium-sized models already lead to model selection challenges.

To date, three approaches for estimating PVAR models exist in the literature. The first two estimate a priori unrestricted PVAR models while the third approach restricts the PVAR model beforehand. The first approach is introduced by Koop and Korobilis (2015b) who propose a Bayesian selection prior, searching for homogeneity and no interdependency restrictions across countries. The second approach is a Bayesian factor approach proposed by Canova and Ciccarelli (2004, 2009). The factor approach reduces the dimension by aggregating information in country-specific, variable-specific and common factors. The third approach sets a priori homogeneity or no dependency restrictions based on theoretical arguments. Canova and Ciccarelli (2013) and Breitung and Roling (2015) show how to estimate these simplified models.

Among others, Banbura et al. (2010), Song and Bickel (2011), Koop and Korobilis (2015b), Nicholson et al. (2016), and Korobilis (2016) provide evidence that using model selection techniques for VAR models which account for specific characteristics such as time series or cross-sectional properties is beneficial in terms of forecasting performance. In line with this, the first two chapters of this thesis introduce two different approaches for finding restrictions for PVAR models, ensuring the estimation feasibility in a data driven way. The proposed selection methods search for restrictions along one or more panel characteristics such as dynamic and static interdependencies and homogeneity among coefficients.

In general, PVAR models are of great interest for empirical analyses. In the unrestricted form they provide an extremely flexible way to model linkages and heterogeneities across multiple variables of several countries. Globally interlinked financial and real markets intensify the interest in spillover analyses. Accounting for an international dimension is crucial, as a model can otherwise be affected by an omitted variable bias due to excluding foreign variables. The application of PVAR models is not restricted to the analysis of dynamics across countries. The models have a wide range of applications as PVAR models can also be used to analyze for example sectoral, firm-specific, or regional data.

However, beside the reduced form estimation, the structural identification of PVAR models is challenging and mainly based on an easy to implement recursive identification. Yet, determining the order of variables in a recursive identification for multi-country models is troublesome due to the lack of theoretical justifications. The third chapter of this thesis applies a three-country VAR model to a spillover analysis and demonstrates a way how to structurally identify shocks in a multi-country model.

In the **first chapter**, *Restrictions Search for Panel Vector Autoregressive Models*, I propose a Bayesian selection prior for PVAR models. The prior differentiates between own countries' and foreign variables. It searches (i) for zero restrictions for coefficients measuring the lagged impact of foreign variables, (ii) for homogeneity across countries among coefficients of domestic variables, and (iii) for correlations between error terms of different countries. The selection prior is a mixture of two normal distributions centering on a restriction with a large and a small variance. The posterior puts more weight on a restriction if it is supported by the data. The findings of a simulation and an empirical application for G7 countries confirm the benefit of the prior. Compared to priors based on less flexible panel structures, models with priors searching for no dynamic interdependency restrictions reduce mean squared errors measuring the deviations of coefficient estimates from the true values in the Monte Carlo simulations and improve forecast performance.

Chapter 1 contributes to the literature on estimation strategies for PVAR models by allowing for a flexible panel structure in the prior specification. It extends the prior of Koop and Korobilis (2015b) by specifying a panel structure which separates domestic and foreign variables. This flexible structure enables one to select the lag length of each variable separately. The posterior probabilities for restrictions provide a ranking based on which restrictions can be set. Furthermore, the prior takes the panel structure into account and thus extends other selection priors for time series data. The flexible panel structure assumed in the prior allows for a wide range of empirical applications being especially beneficial for applications including real and financial variables.

The **second chapter**, *Penalized Estimation of Panel Vector Autoregressive Models: A Lasso Approach*, proposes a new lasso (least absolute shrinkage and selection operator) as an estimation method for PVAR models. This frequentist penalized estimation technique reduces the number of estimated parameters using a penalty term. It sets some coefficients to zero and shrinks others. I introduce a new specification of the penalty term, which builds on characteristics of macroeconomic panel data. That is, first, variables have a decreasing impact on a variable with increasing time distance. Second, variables of one country have a larger impact on variables of the same country than on variables of other countries. To model these properties, the penalty term consists of three parts: First, it constrains coefficients depending on the time distance to the dependent variable. Thus, it captures that more recent variables carry more information for the model. Second, the penalty constrains coefficients from the same country differently than from foreign countries since variables from the same country may be more relevant than variables from a different country. Third, a basic penalty that varies across equations is used. Furthermore, a modified loss function of the optimization problem includes the weighted sum of squared residuals. Thereby, the model does not restrict correlations between error terms of different countries. The results of Monte Carlo simulations and of an empirical forecasting exercise support the use of the lasso for PVAR models. Compared to ordinary least squares estimation, the proposed estimator improves the forecast accuracy, reduces mean squared errors, and is also feasible in very large systems. The lasso for PVAR models outperforms other estimation procedures for PVAR models in large systems.

Chapter 2 contributes to the literature, first, by proposing a new lasso suitable for PVAR models. Second, the asymptotic properties of the lasso for PVAR models are established building on the specification of the penalty term. The estimator asymptotically selects the true variables for inclusion and estimates the nonzero coefficients as efficiently as if the true underlying model were known.

The **third chapter**, *International Monetary Policy Transmission* based on joint work with Gregor von Schweinitz, applies a Bayesian proxy structural three-country VAR model to analyze cross-border effects of monetary policy transmissions. The analysis focuses on international macroeconomic effects of monetary policy spillovers for the United States, the United Kingdom, and the euro area (EA). Conventional and unconventional monetary policy shocks are identified with external instruments. The structural VAR model is augmented with proxy series which embody the effects of central bank announcements on high-frequency forward rates. A Bayesian selection prior which searches for no dynamic interdependencies restrictions on foreign lagged variables is used to estimate the three-country VAR model. The chapter addresses two main questions. First, are there macroeconomic effects of international monetary policy transmission and if so, what do they look like? Second, are there asymmetries across the US, UK, and EA regarding international monetary policy transmission? Our results show no significant cross-border macroeconomic effects of monetary policy surprises.

Chapter 3 contributes to the literature, first, by focusing not only on the effects of US monetary policy transmission but also on the UK and EA. Second, we analyze the international macroeconomic effects of conventional and unconventional monetary policy shocks using a three-country model that allows to capture interlinkages and heterogeneities across the US, UK, and EA. Third, we rely on external instruments for the identification of monetary policy shocks and thus apply an alternative identification to the recursive identification commonly used in multi-country models. The external instrument identification allows us to focus on one shock in a multi-country model without suffering a bias from the neglect of international shock components.

CHAPTER 1

Restrictions Search for Panel Vector Autoregressive Models

1.1 Introduction

Intensifying international goods and knowledge flows, as well as trade agreements, demonstrate the importance of international interdependencies among economies. With these interlinkages empirical analyses require taking both the connections and heterogeneities across countries into account. Recent literature stresses the benefits of including a global dimension while forecasting national and international key macroeconomic variables. Studies using factor models with global factors or multi-country models which account for international linkages provide evidence on improved forecast performance (see e.g., Pesaran et al., 2009; Greenwood-Nimmo et al., 2012; Ciccarelli and Mojon, 2010; Koop and Korobilis, 2015a; Dovern et al., 2016; Garratt et al., 2016; Huber et al., 2016; Bjørnland et al., 2017). Similar, structural spillover analyses disregarding country-specific information and global dependencies could end up with biased results regarding spillover effects and transmission channels (see e.g., Canova and Ciccarelli, 2009; Georgiadis, 2017; Kilian and Lütkepohl, 2017).

One tool that is able to consider dynamic and static global interdependencies as well as cross-sectional heterogeneities is the unrestricted panel vector autoregressive (PVAR) model. A PVAR model includes several countries and country-specific variables in one model. Thus, lagged foreign variables can impact domestic variables, meaning that dynamic interdependencies exist. Static interdependencies between two variables of two countries occur if the covariance between the two is nonzero. Finally, the PVAR model accounts for heterogeneity across countries since the coefficient matrices can vary across economies. This strength of PVAR models comes at the cost of a large number of parameters to estimate - usually set against a relatively low number of time series observations for macroeconomic variables. To overcome this problem, the researcher has to set restrictions on the PVAR model. Typical restrictions for PVAR models are set along interdependencies and homogeneities across countries.

This paper conducts a data-based restrictions search for dynamic and static interdependencies and cross-sectional heterogeneities specifying a Bayesian selection prior for PVAR models. The prior searches (i) for zero restrictions for coefficients measuring the lagged impact of foreign variables, (ii) for homogeneity across countries among coefficients of domestic variables, and (iii) for correlations between error terms of different countries.

The idea of the selection prior is to reduce the dimension of the PVAR model by constraining the parameter space for specific variables. The prior is a mixture of two normal distributions centering around a restriction with a small variance and a large variance. Thus, the first part of the distribution restricts the parameter space while the second part allows for an unrestricted estimation. For which variables the parameter space is restricted, depends on the data. The posterior distribution puts more weight on a restriction if it is supported by the data.

The restricted part of the here proposed prior on lagged foreign coefficients shrinks coefficients to zero, thus, to no dynamic interdependency restrictions. The restricted part of the prior on lagged domestic coefficients shrinks a parameter to the coefficient estimate of another country, thus, to homogeneity restrictions. These prior specifications separate domestic and foreign variables, although foreign variables are not separated on a country basis as priors are set on single lagged coefficients and not on variables grouped for each country. Furthermore, the prior on the covariance matrix of the PVAR model allows for shrinkage to no static interdependencies.

By accounting for panel characteristics in the restrictions search but allowing for a flexible structure, the paper adds to the literature on selection priors. George and McCulloch (1993) develop a selection prior for multiple regression models. Based on a hierarchical prior variables are selected which are included in the model. George et al. (2008) extend the stochastic search variable selection (SSVS) to the use for vector autoregressive (VAR) models. Koop and Korobilis (2015b) develop a selection prior for PVAR models. Their stochastic search specification selection (S^4) builds closely on George et al. (2008) but adds a restrictions search for homogeneity of domestic autoregressive coefficients across countries. Further, in contrast to SSVS, they run the restrictions search on whole matrices including all variables of one country. In order to distinguish the algorithm here proposed from S^4 and SSVS, the algorithm is called stochastic search variable selection for PVAR models (SSVSP).

The panel structure of the SSVSP, separating domestic and foreign variables, is less restrictive than the structure assumed by Koop and Korobilis (2015b). By implementing their prior on country matrices, Koop and Korobilis (2015b) assume a specific panel structure; namely, all variables of one country are treated in a similar way: either restricted or not. The flexible panel structure of the SSVSP has the advantages that, first, the SSVSP can provide evidence supporting the exclusion of a single lag of a variable. A clear ranking of posterior probabilities of which variables to include in the model and which coefficients are homogeneous can be developed. Using the S^4 prior of Koop and Korobilis (2015b) decides on excluding a single variable based on the results for a matrix-wide search. Second, compared to the commonly used Minnesota prior for large Bayesian VAR models, which assumes a specific shrinkage depending on the lag number, the SSVSP takes the panel structure into account. Koop and Korobilis (2015b) as well as Korobilis (2016) provide evidence that a prior for PVAR models which accounts for the inherent panel dimension within the data improves forecast accuracy and reduces mean squared errors. In addition, in a set-up where country grouping for restrictions does not hold, Korobilis (2016) demonstrates that the absolute deviation from the true value is lower for SSVS than it is for S^4. This result contributes to the argument for a restrictions search on single elements. Third, the SSVSP prior has a wider range for empirical applications than does the more rigid S^4. Applications, including financial and real variables, can especially benefit from a less restrictive form since the SSVSP can incorporate variable specific restrictions.

These advantages are reflected in the results of both a Monte Carlo simulation and a forecasting exercise. First, the results of the Monte Carlo studies show that especially when a more flexible panel structure is present, the posterior estimates of the SSVSP deviate less from the true values than the ones of S^4. Furthermore, the SSVSP is accurate in the selection of the restrictions displayed in the posterior probabilities for no interdependencies and homogeneities. Second, the results of the empirical application demonstrate that forecast performance improves for the SSVSP specifications which focus on sparsity in form of no dynamic interdependencies.

In the following, section 1.2 relates the paper to the relevant literature. Section 1.3 describes possible restrictions for PVAR models. Section 1.4 introduces the stochastic search variable selection for PVAR models. Next, in section 1.5 the performance of the SSVSP is evaluated based on two Monte Carlo simulations and in section 1.6 an empirical application is conducted. Finally, section 1.7 concludes.

1.2 Literature

The paper contributes to the Bayesian selection prior literature by providing an estimator which takes the panel dimension into account but allows for a flexible panel structure. The selection prior literature starts with the paper of George and McCulloch (1993), who developed the prior for multiple regression models. The procedure, which the authors call stochastic search variable selection (SSVS), selects the variables that should be included in the regression model. This is achieved by using a hierarchical prior for the coefficients of the right hand side variables. The variables that should be included in the model occur more frequently when sampling from the conditional posterior distributions in the Gibbs sampler. George et al. (2008) further develop the SSVS, extending it for use with VAR models. They set a hierarchical prior on the autoregressive coefficients and find close to zero elements. Additionally, the authors use the prior for structural identification. They decompose the covariance matrix into two upper triangular matrices and let the SSVS algorithm find additional zero restrictions by searching for the elements of the decomposition matrix that are close to zero. Korobilis (2008) and Jochmann et al. (2010) show that forecast performance is improved for VAR models when using the SSVS. The first paper uses the SSVS in a factor model that includes a large number of macroeconomic variables for the United States. The second paper allows for structural breaks. Using data for the United States, the authors show that forecasts improve mainly due to the usage of the SSVS and not due to the consideration of structural breaks. Subsequently, Korobilis (2013) extends the selection priors further to nonlinear set-ups. However, all these papers do not account for a cross-sectional dimension.

Koop and Korobilis (2015b) are the first to develop a selection prior for PVAR models. Their stochastic search specification selection (S^4) builds closely on George et al. (2008) but adds a restrictions search for homogeneity of domestic autoregressive coefficients across countries. Further, in contrast to SSVS, they run the restrictions search on whole matrices including all variables of one country and, thus, assume a specific matrix panel structure. Therefore, the authors call their procedure specification search. Koop and Korobilis (2015b) demonstrate with their Monte Carlo simulation that S^4 estimates are closer to the true values than OLS estimates. Using data for sovereign bond yields, industrial production, and bid-ask spread for euro area countries from January 1999 to December 2012, they show that the model fit improves when taking the characteristics of a panel model into account compared to a BVAR model without restrictions search. Thus, the results of Koop and Korobilis (2015b) provide evidence that accounting for the panel dimension in a prior for PVAR models is beneficial in terms of model fit. Korobilis (2016) comes to the same conclusion. He compares different prior specifications for PVAR models. For larger PVAR models, priors taking the panel dimension into account deviate less from the true values than other VAR priors. In addition, priors with a panel dimension improve the forecasting performance that he demonstrates for the same empirical application as in Koop and Korobilis (2015b). For small samples, however, the Bayesian shrinkage priors cannot outperform the OLS estimates.

A drawback of the prior of Koop and Korobilis (2015b) is the specific country grouping assumed for the restrictions. The grouping constrains statements about interdependencies and heterogeneities on the country level. Thus, differences among variables of a specific country are neglected. Which variables are driving linkages and where homogeneity exist among variables, cannot be assessed. The restrictions search on the matrix-level can lead to exclusions of potentially important variables since all variables of one country are treated in a similar way being restricted or not. Instead, the SSVSP searches for restrictions for each variable and thus can provide evidence supporting the exclusion of a single lag of a variable.

Besides the Bayesian selection prior, as yet, the literature follows two further strands to overcome the curse-of-dimensionality problem in PVAR models: using a Bayesian factor approach and setting restrictions beforehand. My suggested selection prior complements the existing approaches by providing an alternative that overcomes issues of the factor approach and the use of an a priori restricted model.

The Bayesian factor approach is proposed by Canova and Ciccarelli (2004, 2009). Their Bayesian cross-sectional shrinkage approach aggregates the parameters into lower dimensional factors and thereby reduces the number of parameters to estimate. These factors consist of a variable-specific, a country-specific, and a common factor. Canova and Ciccarelli (2012), studying dynamics of the European business cycle, and Ciccarelli et al. (2016), analyzing spillovers in macro-financial linkages across developed economies, apply the cross-sectional shrinkage approach. Billio et al. (2016) extend the approach to a Markov-switching model. Koop and Korobilis (2015a) broaden it to time-varying parameter PVAR models additionally allowing for time-varying covariance matrices. An issue with the cross-sectional shrinkage approach is that the structural identification is more complex since the error term includes two components coming from the equation estimating the factorized parameters and from the estimation of the VAR model. Furthermore, the factors condense the information in the variables. This aggregation can have an impact on the dynamics of the model.

A second way is to assume no dependency and homogeneity across the panel units. Setting a priori assumptions, however, is troublesome as theoretical justification is often insufficient. The assumptions must be based on a solid theoretical background which is not likely to be satisfied in macroeconomic panels. Examples setting assumptions include Love and Zicchino (2006), Gnimassoun and Mignon (2016), and Attinasi and Metelli (2017), assuming homogeneity and no dynamic interdependencies, while

Ciccarelli et al. (2013) or Comunale (2017) restrict for no dynamic interdependencies. Pérez (2015) and Wieladek (2016) use a Bayesian approach and assume no dynamic interdependencies. By setting these restrictions the researcher reduces the number of parameters in the models. Estimation procedures for these kinds of models are described in Canova and Ciccarelli (2013) and Breitung and Roling (2015).

Besides PVAR models, other dynamic times series models which are suitable for modeling a large number of time series accounting for an international dimension are global VAR (GVAR) models, factor augmented VAR (FAVAR) models, and large Bayesian VAR (large BVAR) models. GVAR models introduced by Dees et al. (2007) combine a single country VAR model with international variables constructed as the weighted averages from several countries. The weights depend on a connectivness measure such as trade flows. However, the weights only allow for a rather rigid interdependency structure as weights are usually constant over time and the same for each variable. FAVAR models as introduced by Bernanke et al. (2005) augment a VAR model with latent factors. The factors are extracted from variables which are not included in the VAR model. Given the factor structure the structural identification of these models is challenging. Moreover, the aggregation through the factors can affect the responses of the variables of the model to structural shocks. Large Bayesian VAR models use Bayesian shrinkage as done by Banbura et al. (2010). The standard priors used for large BVAR models do not account for the panel dimension in the data. Thus, these priors are less suited for multi-country analyses.

The paper adds to the literature on using multi-country models for forecasting. As PVAR models allow for international interdependencies, they are excellent tools for forecasting. The suggested prior specification provides a flexible structure which is applicable to a wide range of forecasting applications. Several studies provide evidence that models including a multi-country dimension is beneficial for forecasting. Ciccarelli and Mojon (2010) and Bjørnland et al. (2017) use a factor model for inflation and GDP forecasts, respectively. Koop and Korobilis (2015a) indicate that using a panel VAR, estimated by a factor approach, for forecasting key macroeconomic indicators of Euro zone countries can lead to improvements in forecasts. Pesaran et al. (2009), Greenwood-Nimmo et al. (2012), Dovern et al. (2016), Huber et al. (2016), and Garratt et al. (2016) provide evidence that forecasts based on GVAR models improve forecast performance relative to univariate benchmark models.

1.3 PVAR Model Restrictions

A PVAR model for country i at time t with $i = 1, ..., N$ and $t = 1, ..., T$ is given by

$$y_{it} = A_{i1}Y_{t-1} + A_{i2}Y_{t-2} + ... + A_{iP}Y_{t-P} + u_{it}, \tag{1.1}$$

where $Y_{t-1} = (y'_{1t-1}, ..., y'_{Nt-1})'$ and y_{it} denotes a vector of dimension $[G \times 1]$ where the number of country-specific variables is defined as G.[1] All matrices A_{ip} have dimension $[G \times NG]$ for lag $p = 1, ..., P$. The index i denotes that the matrices are country specific for country i. The u_{it} are uncorrelated over time and normally distributed with mean zero and covariance matrix Σ_{ii}. The assumption of normally distributed error terms is used for deriving the posterior distributions as the form of the likelihood function is determined by the normality assumption. The covariance matrix between errors of different countries is defined as $E(u_{it}u'_{jt}) = \Sigma_{ij} \quad \forall\ i \neq j$ with dimension $[G \times G]$.

The PVAR model for all N countries can then be written as

$$Y_t = A_1Y_{t-1} + A_2Y_{t-2} + ... + A_PY_{t-P} + U_t. \tag{1.2}$$

The Y_t and U_t are $[NG \times 1]$-vectors. The U_t is normally distributed with mean zero and covariance matrix Σ that is of dimension $[NG \times NG]$. The $[NG \times NG]$-matrix A_p for one particular lag p is defined as

$$A_p = \begin{pmatrix} \alpha^{11}_{p,11} & \cdots & \alpha^{1k}_{p,1j} & \cdots & \alpha^{1G}_{p,1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha^{l1}_{p,i1} & \cdots & \alpha^{lk}_{p,ij} & \cdots & \alpha^{lG}_{p,iN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \alpha^{G1}_{p,N1} & \cdots & \alpha^{Gk}_{p,Nj} & \cdots & \alpha^{GG}_{p,NN} \end{pmatrix}.$$

The element $\alpha^{lk}_{p,ij}$ refers to the coefficient of lag p of variable k of country j in the equation of variable l of country i. Thus, it measures the impact of lag p of variable k of country j on variable l of country i.

A structural form of the PVAR model is derived by decomposing the covariance matrix Σ into $\Sigma = \Psi^{-1'}\Psi^{-1}$ where Ψ is an upper triangular matrix. Therefore, the structural identification is based on a recursive order. An element ψ^{lk}_{ij} of the upper triangular matrix Ψ defines the static relation between variable l of country i and variable k of country j.

This structural PVAR model can account for dynamic interdependencies (DI), static

[1]Although this specification does not include a constant, it can be extended to include one.

interdependencies (SI), and cross-sectional heterogeneities (CSH).[2] By allowing for these interdependencies and heterogeneities, unrestricted PVAR models provide an extremely flexible way to model linkages and heterogeneities across multiple variables of several countries. However, the large number of free parameters in the unrestricted PVAR model requires some kind of model selection. A straightforward way of reducing the dimension of a PVAR model is imposing restrictions using the panel structure inherent in the data. It can be expected that interdependencies and heterogeneities across countries only exist for specific country and variable combinations. Therefore, for some coefficients the following restrictions can hold:

1. **No dynamic interdependencies (DI)**: no lagged impact from variable l of country i to variable k of country j for lag p if $\alpha_{p,ij}^{lk} = 0$ for $j \neq i$.

2. **No static interdependencies (SI)**: no correlation between the error term of equation l of country i, u_{it}^{l}, with the error term of equation k of country j, u_{jt}^{k}, if $\psi_{ij}^{lk} = 0$ for $j \neq i$.

3. **No cross-sectional heterogeneities (CSH)**: homogeneous coefficient across the economies for lag p if $\alpha_{p,jj}^{lk} = \alpha_{p,ii}^{lk}$ for $j \neq i$.

In total, $[(NG - G)NG]$ dynamic interdependencies, $[(N(N-1)/2)G^2]$ static interdependencies, and $[(N(N-1)/2)G^2]$ cross-sectional heterogeneities restrictions can be defined.[3] The essential part is to determine for which country and variable combinations these restrictions hold.

1.4 Selection Prior for PVAR Models

The unrestricted PVAR model with one lag can be rewritten as

$$Y_t = Z_{t-1}\alpha + U_t, \tag{1.3}$$

where α is the vectorized matrix A and $Z_{t-1} = (I_{NG} \otimes Y_{t-1})$. The model is simplified to a model including only one lag.[4] Otherwise, the restrictions search for dynamic interdependencies would provide guidance for lag length selection.

Each element of α is drawn from a mixture of two normal distributions centering around the restriction, either with a small or large variance. The coefficient either shrinks to the restriction (small variance case) or is estimated with a looser prior

[2] Canova and Ciccarelli (2013) provide a survey of the PVAR model restrictions.

[3] Note that while SI restrictions are symmetric, this is not (necessarily) the case for DI restrictions.

[4] Focusing on the model with one lag allows to drop the lag index p from the notation. Thus, the coefficient α_{ij}^{lk} is the coefficient for $p = 1$.

(larger variance case). Thus, the algorithm imposes soft restrictions by allowing for a small variance. In contrast to Koop and Korobilis (2015b), the restrictions search is completed for each single element and not on whole matrices that include all variables for a given country.

The SSVSP algorithm has specific priors for the parameters of A and for the covariance matrix building on the DI, SI, and CSH restrictions. The DI restrictions impose limits on the coefficients of the lagged foreign endogenous variables. The DI prior is given by

$$\begin{aligned} \alpha_{ij}^{lk} \mid \gamma_{DI,ij}^{lk} &\sim (1-\gamma_{DI,ij}^{lk})\mathcal{N}(0,\tau_1^2) + \gamma_{DI,ij}^{lk}\mathcal{N}(0,\tau_2^2) \\ \gamma_{DI,ij}^{lk} &\sim Bernoulli(\pi_{DI,ij}^{lk}). \end{aligned}$$

The prior distribution of α_{ij}^{lk} is conditional on the hyperparameter $\gamma_{DI,ij}^{lk}$ which is set to be Bernoulli distributed. If $\gamma_{DI,ij}^{lk}$ is equal to zero, α_{ij}^{lk} is drawn from the first part of the normal distribution with mean zero and variance τ_1^2.[5] If $\gamma_{DI,ij}^{lk}$ is equal to one, α_{ij}^{lk} is drawn from the second part of the normal distribution with mean zero and variance τ_2^2. The values of τ_1^2 and τ_2^2 must be chosen such that τ_1^2 is smaller than τ_2^2. Thus, if $\gamma_{DI,ij}^{lk} = 0$, the prior is tight in the sense that the parameter is shrunk to zero and no dynamic interdependency is supported by the data. Whereas the prior is loose for $\gamma_{DI,ij}^{lk} = 1$.

The SI prior is set on the elements of the upper triangular matrix, Ψ. If SI restrictions are found, the structural PVAR model is overidentified since additional zero restrictions can be set on top of the recursive ordering. A clear advantage of this decomposition is that it assures that by construction every simulated Σ is positive definite.[6]

The prior for SI restrictions follows the same logic as the DI prior:

$$\begin{aligned} \psi_{ij}^{lk} \mid \gamma_{SI,ij}^{lk} &\sim (1-\gamma_{SI,ij}^{lk})\mathcal{N}(0,\kappa_1^2) + \gamma_{SI,ij}^{lk}\mathcal{N}(0,\kappa_2^2) \\ \gamma_{SI,ij}^{lk} &\sim Bernoulli(\pi_{SI,ij}^{lk}). \end{aligned}$$

The prior is set for all $j \neq i$. To assure positive variance elements, the $(\psi_{ii}^{kk})^2$ are gamma distributed, $(\psi_{ii}^{kk})^2 \sim \mathcal{G}(a,b)$. The elements for the same country, ψ_{ii}^{lk} for $l \neq k$, are normally distributed with mean zero and variance κ_2^2. All ψ_{ij}^{lk} elements $(j \neq i)$ are drawn from the specified hierarchical prior. The parameter κ_1^2 is smaller than κ_2^2. If $\gamma_{SI,ij}^{lk}$ is equal to zero, the parameter shrinks to zero showing that the data do not support static interdependency. The selection prior can be used to estimate the reduced form of a PVAR model by not applying a Cholesky decomposition and not searching

[5]How $\pi_{DI,ij}^{lk}$ is set is described in detail in 1.B. This holds also for the CSH and SI priors.
[6]Compare also to Koop and Korobilis (2015b).

for static interdependency restrictions but by assuming a standard distribution for the PVAR model variance, e.g. an inverse Wishart distribution.

Searching for homogeneity across countries is not as straightforward as searching for the zero restrictions for dynamic and static interdependencies. The main contribution of Koop and Korobilis (2015b) is the development of a procedure how to search for CSH restrictions. Possible homogeneity across countries is assessed for the coefficients measuring the impact of domestic variables on variables of the same country. The CSH prior is given by

$$\begin{aligned}\alpha_{jj}^{lk} \mid \gamma_{CSH}^{w} &\sim (1-\gamma_{CSH}^{w})\mathcal{N}(\alpha_{ii}^{lk}, \xi_1^2) + \gamma_{CSH}^{w}\mathcal{N}(0, \xi_2^2)\\ \gamma_{CSH}^{w} &\sim Bernoulli(\pi_{CSH}^{w}).\end{aligned}$$

The prior is for all $j \neq i$. There are $(N(N-1)/2)G^2 = K$ combinations of coefficients that are checked for homogeneity. The index $w = 1, ..., K$ refers to a specific combination. Again, ξ_1^2 is smaller than ξ_2^2. The main difference to the DI and SI prior is that instead of shrinking the parameter to zero in the first part of the normal distribution, the mean is equal to the coefficient for which homogeneity is being checked, $\alpha_{jj}^{lk} = \alpha_{ii}^{lk}$ in mean.

To be able to check all possible combinations, the procedure of Koop and Korobilis (2015b) is followed, who define a selection matrix $\Gamma = \prod_{w=1}^{K} \Gamma_w$. The matrix Γ_w is an identity matrix of dimension $[NG \times NG]$ with two exceptions. The diagonal element at the position α_{jj}^{lk} is set equal to γ_{CSH}^{w} and the off-diagonal element referring to the element α_{ii}^{lk} is set equal to $(1-\gamma_{CSH}^{w})$. If all coefficients are heterogeneous, all Γ_w are identity matrices. To impose the CSH restrictions, the posterior mean of α is multiplied by the selection matrix Γ.

The prior specification is based on the assumption of a stationary VAR model. The prior shrinks the coefficient estimates to zero. If stochastic trends are present in the data, a prior accounting for integrated variables can shrink coefficient estimates to a multivariate random walk as suggested in the Minnesota prior of Litterman (1986) and Doan et al. (1984). Such an extension could be easily implemented in the model as the unrestricted part of the normal distributions can be set up in a Minnesota prior way. If cointegration relations exist, the use of vector error correction models with appropriate priors is more suitable (for a survey on Bayesian estimation of vector error correction models see Koop et al., 2005; Karlsson, 2013).

To summarize, consider a simple 3-countries-2-variables example, where the variables are ordered such that the first two equations model the dynamics of the two variables of the first country, equations three and four of the second country and the last two equations of the third country. The coefficients of A, which are checked for dynamic interdependencies, are marked with DI and the coefficients checked for homogeneity are marked with CSH. The elements of the covariance matrix which are checked for static interdependencies are marked with SI:

$$A = \begin{pmatrix} CSH & CSH & DI & DI & DI & DI \\ CSH & CSH & DI & DI & DI & DI \\ DI & DI & CSH & CSH & DI & DI \\ DI & DI & CSH & CSH & DI & DI \\ DI & DI & DI & DI & CSH & CSH \\ DI & DI & DI & DI & CSH & CSH \end{pmatrix}, \Psi = \begin{pmatrix} & & SI & SI & SI & SI \\ & & SI & SI & SI & SI \\ & & & & SI & SI \\ & & & & SI & SI \\ & & & & & \\ & & & & & \end{pmatrix}.$$

The prior specification has the advantage that it allows the usage of the Gibbs sampler to sample from the posterior distributions.[7] The means of the posterior distributions are used as point estimates for the coefficients.

The outcome of the algorithm can be interpreted in two ways: model selection and Bayesian model averaging.[8] The researcher can select one specific restricted PVAR model based on the posterior means of γ_{DI}, γ_{SI}, and γ_{CSH}. The researcher can set the restrictions successively until the model with the best fit is found. In addition, a threshold value can be used, often set to 0.5 by the selection prior literature. If the posterior mean of γ is below the threshold value, the restriction is set. Alternatively, the outcome of the algorithm can be used as a Bayesian model averaging (BMA) result. Thus, the posterior means averaged over all draws are taken as coefficient estimates. Since each draw leads to a specific restricted model, the BMA results average over all possible restricted models.

One problematic issue is that the selection prior requires the SUR form of a VAR model, leading to the inversion of large matrices. This leads to a computationally demanding algorithm for medium and large size VAR models.[9] To overcome this problem, Koop (2013) develops a natural conjugate selection prior for VAR models. Here, no MCMC methods must be used. However, the natural conjugate selection prior has two disadvantages.[10] First, each variable can only be either included or excluded in the whole VAR system. Second, the natural conjugate specification requires a specific covariance prior. Thus, a restrictions search for the covariance elements of the VAR model is not possible. Hence, for the purpose, being able to include static interdepen-

[7] The Gibbs sampler algorithm is described in detail in 1.A.

[8] Compare to the general survey in Koop and Korobilis (2010) or the specific explanation for the S^4 in Koop and Korobilis (2015b).

[9] Both Koop (2013) and Korobilis (2013) elaborate further on this issue.

[10] Koop (2013) explains the disadvantages of the natural conjugate prior in detail.

dencies and to allow for dynamic interdependencies that are not homogeneous across countries, the natural conjugate SSVS prior is not an alternative. Instead, the computational burden is accepted for having a differentiated prior that is able to account for the characteristics of a PVAR model, which should be less of a problem with increasing computational capacities.

1.5 Monte Carlo Simulation

1.5.1 Simulation Set-Ups

In order to evaluate the prior, two Monte Carlo simulations are conducted.[11] For both Monte Carlo simulations data are generated from a PVAR model which includes three countries, two variables for each country, one lag, and 100 observations. First, it is assumed that both dynamic and static interdependencies as well as cross-sectional heterogeneities exist for specific variable and country combinations. In particular, country 2 has a dynamic impact on country 1 and country 1 on country 3. Country 3 does not impact the other two countries dynamically. Coefficients are homogeneous between countries 2 and 3. Static interdependencies exist between country 1 and 2. This example has a clear country grouping structure. A scenario like this is given by the first Monte Carlo simulation where the following parameter values are set:

$$A^{true} = \begin{pmatrix} 0.8 & 0 & 0.2 & 0.2 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0.3 & 0 & 0 \\ 0 & 0 & 0.6 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0.3 & -0.4 & 0 & 0 & 0.6 & 0.5 \\ 0.2 & 0.4 & 0 & 0 & 0 & 0.5 \end{pmatrix}, \Psi^{true} = \begin{pmatrix} 1 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 1 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Second, no clear country grouping for interdependencies and homogeneities is assumed. Hence, a less restrictive panel structure exists. The second Monte Carlo simu-

[11] 100 samples, each with a length of 100 are simulated. The Gibbs sampler is done with 55000 draws, of which 5000 draws are disregarded as draws of the burn-in-phase. The calculation is based on a further development of the MATLAB code provided by Koop and Korobilis (2015b) (https://sites.google.com/site/dimitriskorobilis/matlab/panel_var_restrictions).

lation incorporates these properties and has the following true parameters:

$$A^{true} = \begin{pmatrix} 0.8 & 0 & 0.2 & 0 & 0.2 & 0 \\ 0 & 0.7 & 0.2 & 0 & 0.2 & 0 \\ 0 & 0 & 0.6 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0 & 0 \\ 0.3 & -0.4 & 0 & 0 & 0.6 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{pmatrix}, \Psi^{true} = \begin{pmatrix} 1 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The performance of the SSVSP is compared to the performance of different prior specifications. A model with no restrictions search is used as the benchmark mark model, referred to as unrestricted VAR model. This model is estimated using the SSVSP prior but with fixed γ values such that each parameter is drawn from the distribution with the larger variance. Furthermore, the SSVSP is compared to the S^4 and to the SSVS of George et al. (2008).[12] The SSVS prior sets the DI prior on all lagged values and the SI prior on all covariance elements. Thus, no distinction between domestic and foreign variables is made. Additionally, two specifications of the SSVSP are analyzed. First, the SSVSP only searches for DI restrictions (abbreviated with SSVSP_DI), meaning that the γ values for the SI and CSH priors are set to one. Second, the restrictions search is only done for CSH restrictions (SSVSP_CSH).

The performance of each estimator is checked via the absolute percentage deviation (APD) statistic and mean squared errors (MSE):

$$APD = \frac{1}{(NG)^2} \sum_{i=1}^{(NG)^2} \mid \alpha_i - \alpha_i^{true} \mid \quad \text{and} \quad MSE = \frac{1}{(NG)^2} \sum_{i=1}^{(NG)^2} (\alpha_i - \alpha_i^{true})^2.$$

APD (MSE) measures the absolute (squared) deviation of the estimated coefficient α_i, given by the posterior mean averaged over all simulation draws, from the true value, α_i^{true}.[13] The same statistics are computed for Σ. Furthermore, the accuracy of the SSVSP to find the restrictions is evaluated. The posterior probabilities that $\alpha_{ij}^{lk} = 0$, $\psi_{ij}^{lk} = 0$, and $\alpha_{jj}^{lk} = \alpha_{ii}^{lk}$ im mean are compared among themselves and in relation to the true values. These posterior probabilities are calculated as the proportion of $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w} draws that equal zero averaged over all Gibbs sampler draws and all simulated samples. The higher the proportion of γ draws that equal zero, the

[12]The prior hyperparameters used in the Monte Carlo simulations for all different priors are 1.1.

[13]Koop and Korobilis (2015b) uses the absolute deviation and Korobilis (2016) use the mean absolute deviation statistics to evaluate the performance of estimators in Monte Carlo simulations.

Table 1.1: Deviation for estimated coefficient matrix A and Σ from the true values

	Simulation 1				Simulation 2			
	A		Σ		A		Σ	
	APD	MSE	APD	MSE	APD	MSE	APD	MSE
SSVSP	0.043	0.004	0.163	0.042	0.036	0.004	0.085	0.026
SSVSP_DI	**0.026**	**0.001**	0.142	0.030	**0.023**	**0.002**	0.074	0.020
SSVSP_CSH	0.041	0.004	0.182	0.051	0.037	0.004	0.088	0.028
S^4	0.048	0.010	0.109	0.068	0.056	0.011	0.080	0.022
SSVS	0.067	0.008	**0.060**	**0.007**	0.066	0.009	**0.052**	**0.006**
unrest VAR	0.027	**0.001**	0.109	0.020	0.027	**0.002**	0.079	0.020

Note: Absolute (APD) and squared (MSE) deviation of the estimates from the true value for coefficient matrix A and covariance Σ, average over 100 MC draws and all coefficients. Coefficient estimates are the posterior means averaged over all MC draws. Lowest values for each column are in bold. SSVSP_DI: SSVSP with only DI restrictions. SSVSP_CSH: SSVSP with only CSH restrictions. S^4: prior of Koop and Korobilis (2015b). SSVS: prior of George et al. (2008). Unrest VAR: parameters drawn from unrestricted part of distributions. Simulation 1: DGP has matrix panel structure. Simulation 2: DGP has flexible panel structure.

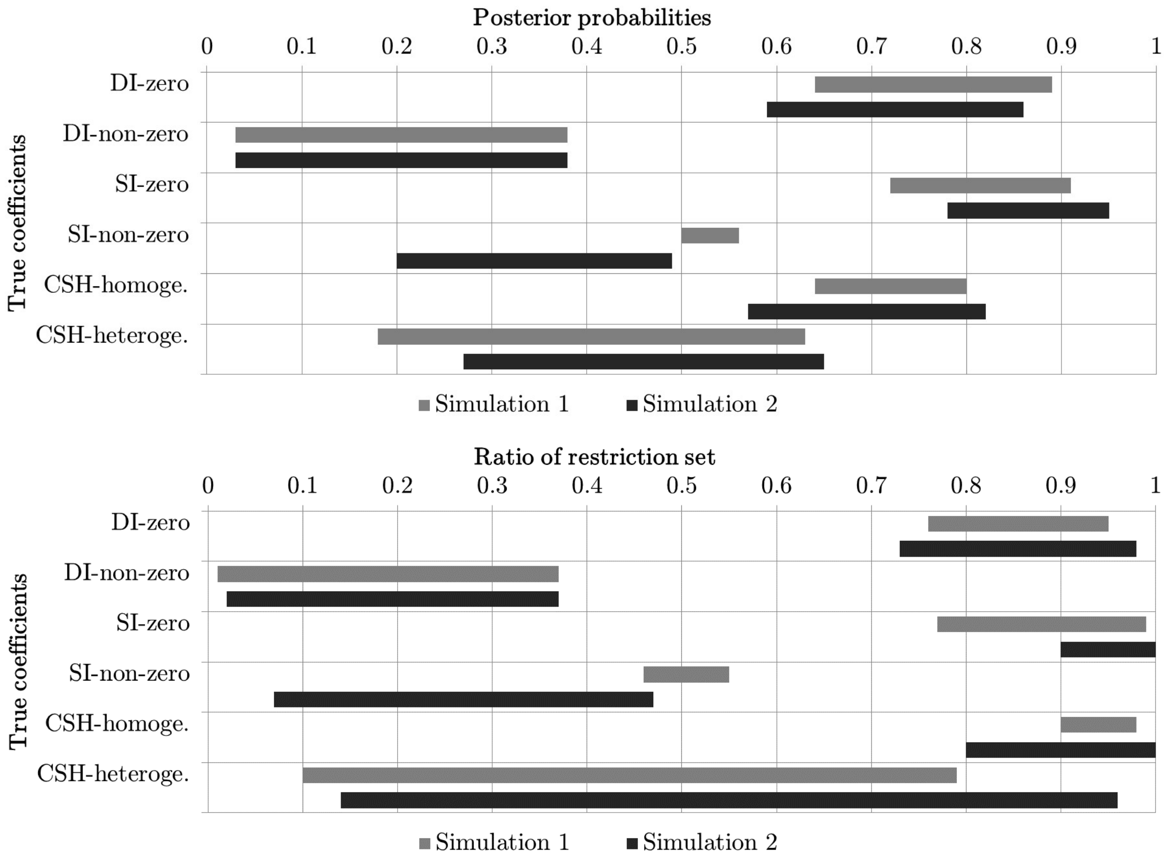
higher is the probability that no dynamic and no static interdependencies exist and coefficients are homogeneous.

1.5.2 Results

The results of the Monte Carlo study demonstrate that, first, the SSVSP outperforms the S^4 in terms of closeness to the true coefficients. This especially holds when a less restrictive panel structure is present. Second, the SSVSP accurately selects the restrictions for the DGPs of both simulations. This is validated by the higher posterior probabilities for no interdependencies and homogeneity for parameters which are truly zero or homogeneous compared to the probabilities for nonzero and heterogeneous parameters.

As table 1.1 shows, the estimated coefficients which are the posterior means averaged over all simulation draws from the SSVSP are on average slightly closer to the true values compared to S^4 for both simulations. In particular, the S^4 performs weaker in simulation 2, where a less restrictive panel structure is present, since the grouping structure on which the restrictions search is done is not present in the data. Furthermore, the APD and MSE of A favor SSVSP_DI for both simulations with $APD_{SSVSP\_DI} = 0.026$ and $MSE_{SSVSP\_DI} = 0.001$ for simulation 1 and $APD_{SSVSP\_DI} = 0.023$ and $MSE_{SSVSP\_DI} = 0.002$ for simulation 2. However, the unrestricted VAR model outperforms the SSVSP and S^4. Doing the restrictions search only for CSH reduces the average deviation from the true values only in simulation 1

Figure 1.1: Range of posterior probabilities $\alpha_{ij}^{lk} = 0$, $\psi_{ij}^{lk} = 0$, and $\alpha_{jj}^{lk} = \alpha_{ii}^{lk}$ and ratio of restriction set with threshold value 0.5



Note: Posterior probabilities, $p(\alpha_{ij}^{lk} = 0)$, $p(\psi_{ij}^{lk} = 0)$, and $p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$, are calculated as the proportion of $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w} draws that equal zero averaged over all Gibbs sampler draws and all simulated samples. Ratios are calculated as the number of restrictions set with threshold 0.5 averaged over all Gibbs draws for each simulated draw relative to all simulated draws. True coefficients are the parameters set when simulating the DGPs. Simulation 1: DGP has matrix panel structure. Simulation 2: DGP has flexible panel structure.

compared to the SSVSP. This indicates that the gain is not explained by the reduced number of restrictions on which the search is done for but rather by searching for no dynamic interdependencies. Thus, the use of a prior which incorporates no dynamic interdependencies is beneficial for the DGPs of both simulations.

For Σ the lowest values for APD and MSE are obtained with the SSVS. In general, deviations from the true values are higher for Σ compared to A. This could be a result of using a symmetric distribution rather than the usually preferred inverse Wishart distribution for covariances.

Furthermore, the SSVSP algorithm is accurate in selecting the restrictions. Posterior probabilities that no interdependencies or heterogeneities exist are higher for true zero or homogeneous values compared to nonzero or heterogeneous values, shown in the

Table 1.2: Accuracy of selecting DI restrictions: posterior probabilities $p(\alpha_{ij}^{lk} = 0)$

Simulation 1						Simulation 2					
-	-	0.2	0.2	0	0	-	-	0.2	0	0.2	0
		0.26	*0.38*	*0.88*	*0.78*			*0.30*	*0.62*	*0.38*	*0.59*
-	-	0.3	0.3	0	0	-	-	0.2	0	0.2	0
		0.06	*0.18*	*0.83*	*0.76*			*0.31*	*0.70*	*0.30*	*0.70*
0	0	-	-	0	0	0	0	-	-	0	0
0.71	*0.75*			*0.85*	*0.73*	*0.82*	*0.73*			*0.84*	*0.65*
0	0	-	-	0	0	0	0	-	-	0	0
0.74	*0.77*			*0.89*	*0.79*	*0.84*	*0.75*			*0.86*	*0.71*
0.3	-0.4	0	0	-	-	0.3	-0.4	0	0	-	-
0.15	*0.03*	*0.79*	*0.64*			*0.19*	*0.03*	*0.76*	*0.67*		
0.2	0.4	0	0	-	-	0	0	0	0	-	-
0.25	*0.05*	*0.78*	*0.66*			*0.86*	*0.79*	*0.83*	*0.71*		

Table 1.3: Accuracy of selecting SI restrictions: posterior probabilities $p(\psi_{ij}^{lk} = 0)$

Simulation 1						Simulation 2					
-	-	0.5	0.5	0	0	-	-	-0.5	0	0	0
		0.56	*0.54*	*0.72*	*0.80*			*0.49*	*0.95*	*0.78*	*0.95*
-	-	-0.5	-0.5	0	0	-	-	0	-0.5	0	0
		0.56	*0.50*	*0.74*	*0.76*			*0.91*	*0.20*	*0.89*	*0.91*
-	-	-	-	0	0	-	-	-	-	0	0
				0.84	*0.86*					*0.90*	*0.93*
-	-	-	-	0	0	-	-	-	-	0	0
				0.91	*0.91*					*0.92*	*0.93*

Note: Posterior probabilities, $p(\alpha_{ij}^{lk} = 0)$ and $p(\psi_{ij}^{lk} = 0)$ in italic, are calculated as the proportion of $\gamma_{DI,ij}^{lk}$ or rather $\gamma_{SI,ij}^{lk}$ draws that equal zero averaged over all Gibbs sampler draws and all simulated samples. Simulation 1: DGP has matrix panel structure. Simulation 2: DGP has flexible panel structure.

first graph of figure 1.1. The first graph presents the range of posterior probabilities for simulations 1 and 2 for true zero or homogeneous coefficients and true nonzero or true heterogeneous coefficients. The second graph demonstrates the range of the ratios of restriction set when doing model selection. The ratios are calculated as the number of restriction set for each MC draw relative to all simulated draws given a threshold value of 0.5.[14] The ratio of restrictions set is higher for true zero and homogeneous values than for nonzero and somehow for heterogeneous values. Thus, the first graph demonstrates the accuracy of selecting restrictions in case of BMA while the second graph underlines it for model selection.

[14]The share of restrictions set for threshold values 0.6, 0.7, 0.8, and 0.9 for simulation 1 and 2 are given in 1.C.

Table 1.4: Accuracy of selecting CSH restrictions: posterior probabilities $p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$

	Simulation 1			Simulation 2		
coefficients	true α_{jj}^{lk}	true α_{ii}^{lk}	$p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$	true α_{jj}^{lk}	true α_{ii}^{lk}	$p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$
α_{11}^{11} & α_{22}^{11}	0.8	0.6	*0.61*	0.8	0.6	*0.65*
α_{11}^{21} & α_{22}^{21}	0	0	*0.73*	0	0	*0.79*
α_{11}^{12} & α_{22}^{12}	0	0.5	*0.19*	0	0.5	*0.28*
α_{11}^{22} & α_{22}^{22}	0.7	0.5	*0.54*	0.7	0.3	*0.30*
α_{11}^{11} & α_{33}^{11}	0.8	0.6	*0.63*	0.8	0.6	*0.64*
α_{11}^{21} & α_{33}^{21}	0	0	*0.75*	0	0	*0.82*
α_{11}^{12} & α_{33}^{12}	0	0.5	*0.18*	0	0.5	*0.27*
α_{11}^{22} & α_{33}^{22}	0.7	0.5	*0.57*	0.7	0.5	*0.40*
α_{22}^{11} & α_{33}^{11}	0.6	0.6	*0.79*	0.6	0.6	*0.75*
α_{22}^{21} & α_{33}^{21}	0	0	*0.80*	0	0	*0.80*
α_{22}^{12} & α_{33}^{12}	0.5	0.5	*0.64*	0.5	0.5	*0.57*
α_{22}^{22} & α_{33}^{22}	0.5	0.5	*0.65*	0.3	0.5	*0.48*

Note: Posterior probabilities, $p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$ in italic, are calculated as the proportion of γ_{CSH}^{w} draws that equal zero averaged over all Gibbs sampler draws and all simulated samples. Simulation 1: DGP has matrix panel structure. Simulation 2: DGP has flexible panel structure.

In detail, looking at simulation 1 and DI restrictions, the probabilities that $\alpha_{ij}^{lk} = 0$ are considerably higher for true zero parameters than for true nonzero values. The first are in a range between 0.64 and 0.89 while the latter one are between 0.03 and 0.38. Table 1.2 provides the detailed results for the posterior probabilities for $\alpha_{ij}^{lk} = 0$. Turning to simulation 2, if no dynamic interdependencies occur in truth, the probabilities that $\alpha_{ij}^{lk} = 0$ are between 0.59 and 0.86. Thus, they are clearly higher than the probabilities for the parameters which dynamically affect the dependent variables, between 0.03 and 0.38. Findings are in the same range for the ratio of restrictions set.

The SSVPS selects accurately the SI restrictions in both simulations as shown in figure 1.1 and table 1.3. This is true since for both simulations the probabilities that $\psi_{ij}^{lk} = 0$ and ratios of restriction set are higher for true zero compared to nonzero parameters. The results for simulation 1 show that probabilities are in a range of 0.72 and 0.91 for zero values while for the existing static interdependencies probabilities are between 0.50 and 0.56. For simulation 2 the probabilities for no static interdependencies, between 0.78 and 0.95, are clearly higher for the true zero values compared to the probabilities for nonzero values, 0.20 and 0.49.

Moreover, the SSVSP is mostly accurate in the selection of the cross-sectional heterogeneity restrictions. The detailed results for $p(\alpha_{jj}^{lk} = \alpha_{ii}^{lk})$ are presented in table 1.4. For both simulations probabilities that the coefficients are homogeneous are higher for true homogeneous coefficients (with few exceptions for simulation 2). However, especially for true values which are close to each other but not equal, the probabilities for homogeneity are relatively high with values above 0.5. Again, the same holds for the ratio of restrictions set. This slightly weaker performance of the SSVSP to pick the correct CSH restrictions compared to DI and SI was already visible in the APD and MSE results for the SSVSP_CSH.

1.6 Empirical Application

1.6.1 Data and Procedure

Using an empirical application as an example, the SSVSP is validated based on its forecasting performance, on the restriction posterior probabilities, and on an impulse response analysis. The PVAR(1) model includes three key macroeconomic variables for the G7 countries - a growth rate of industrial production (IP), a CPI growth rate (CPI), and a short term interest rate (IR). The countries are Canada (CA), Italy (IT), United Kingdom (UK), France (FR), Japan (JP), Germany (DE), and the United States (US). The data are from the OECD and have monthly frequency from 1990:1 through 2015:2.[15] The model which is considered here serves as an illustration for the performance of the SSVSP. In many ways it is not the best model for the DGP as it, for example, takes into account only a fraction of variables which could be of interest for assessing the question of global spillovers or conducting forecasts. Furthermore, the lag order one is set by assumption and is not further validated.

At first, forecasts are provided for 12 horizons for the period beginning from January 2005 through the end of the sample.[16] To obtain the forecasts a predictive distribution is simulated based on the reduced form of the PVAR model with normally distributed error terms. The reduced form model is the model where no SI restrictions search is done and the covariance matrix is drawn from an inverse Wishart distribution.[17] The forecasts are evaluated using the mean squared forecast error (MSFE) and the average predictive likelihoods (PL). The MSFE is calculated as the difference between the estimated forecast, which is given by the posterior mean of the predictive distribution, and the true value given by the data. The MSFEs for a specific variable and horizon are averaged over all forecasts. The PL is the posterior predictive density evaluated at

[15] The hyperparameters of the prior distributions are set as in the Monte Carlo simulations. Detailed information is provided in 1.B. 110,000 draws are computed for the Gibbs sampler, the first 10,000 are disregarded as burn-in-phase.

[16] The forecasts for the included 21 variables are generated iteratively. Forecasts start conditional on the data from January 1990 to December 2004.

[17] The covariance matrix is drawn from an inverse Wishart distribution with T degrees of freedom and identity matrix plus sum of squared residuals as scaling matrix.

the true observation y_{t+h}.

The forecast performance is compared to the unrestricted VAR model, SSVSP_DI, SSVSP_CSH, S^4, and SSVS. Furthermore, two specifications are added which access the selection property of the SSVSP: SSVSP_setDI_v1 and SSVSP_setDI_v2. The SSVSP_setDI_v1 uses the outcome of the SSVSP and sets zeros whenever the posterior probability for a DI restriction is larger than 0.99. The SSVSP_setDI_v2 uses 0.5 as a threshold value.

For the structural analysis the variables are ordered in a recursive way. The industrial production growth rate is ordered first, CPI growth rate second, and the short term interest rate third. The monetary policy shock for one country is thus identified by the assumption that the interest rate does not react contemporaneously to unexpected changes in real variables while a monetary policy shock instantaneously impacts the two real variables. The recursive country ordering is based on the openness of a country. Openness is measured based on yearly import and export data for the economies. The higher the trade of a country is, the more open it is. The countries are arranged in ascending order meaning that the most open country, the United States, is ordered last. Thus, US variables can influence all other countries contemporaneously but are not affected by the variables of the remaining G7 countries.

1.6.2 Results

The results of the empirical application demonstrate three key findings. First, the MSFEs and PLs results favor the SSVSP_DI and the two selection models, SSVSP_setDI_v1 and SSVSP_setDI_v2, indicating that restrictions search is beneficial since sparsity in form of no dynamic interdependencies exist. However, the very large number of restrictions searched for in the SSVSP leads to relatively weak forecast performance. Second, the posterior probabilities for the restrictions indicate that domestic interest rates and inflation evolve unaffected by lagged foreign industrial production growth rates, validated by high posterior probabilities for no dynamic interdependencies. The interest rate of a country depends likely statically and dynamically on foreign interest rates. No heterogeneities are in particular found for the effect of domestic industrial production growth on the domestic interest rate and inflation. Third, the impulse response analysis supports the reliability of the results. In the following the key findings are explained in more detail.

Table 1.5 shows in percent the number of MSFEs which are smaller or equal to one averaged over all variables for forecast horizons 1, 2, 4, 6, and 12. The MSFEs are given relative to the unrestricted VAR model. Thus, a MSFE smaller than one indicates improved forecast performance relative to the unrestricted VAR model. Between

Table 1.5: Mean squared forecast errors relative to unrestricted VAR model

	number of MSFEs $\leq$ 1(in %)				
	horizon 1	horizon 2	horizon 4	horizon 6	horizon 12
SSVSP	38.10	42.86	33.33	28.57	33.33
SSVSP_DI	100.00	85.71	85.71	95.24	90.48
SSVSP_CSH	33.33	38.10	19.05	23.81	52.38
SSVSP_setDI_v1	0.00	0.00	52.38	57.14	66.67
SSVSP_setDI_v2	4.76	4.76	52.38	71.43	80.95
S^4	42.86	33.33	38.10	33.33	47.62
SSVS	19.05	9.52	66.67	95.24	76.19

Note: MSFE relative to unrestricted VAR model. Forecasts for 12 horizons for Jan 2005 to end of the sample. Unrestricted VAR model: parameters drawn from unrestricted part of distributions. S^4: prior of Koop and Korobilis (2015b). SSVS: prior of George et al. (2008). SSVSP_DI: SSVSP with only DI restrictions. SSVSP_CSH: SSVSP with only CSH restrictions. SSVSP_setDI_v1: threshold 0.99 to set zero DI restrictions. SSVSP_setDI_v2: threshold 0.5 to set zero DI restrictions. Σ drawn from inverse Wishart distribution with T degrees of freedom and identity matrix plus sum of squared residuals as scaling matrix. 110,000 Gibbs draws, 10,000 disregarded as burn-in-phase.

28.57% and 42.86% of the MSFEs of the SSVSP are below or equal to the MSFEs of the unrestricted VAR model. Thus, the SSVSP cannot improve the forecasts compared to the unrestricted VAR model. This is quite similar to the performance of the S^4.

The SSVSP_DI performs particularly well. It outperforms the unrestricted model at the best in 100.00% of the cases (horizon 1) and at worst in 85.71% of the cases (horizons 2 and 4). Since the number of restrictions which are examined in the SSVSP are high, the information in the data might not be enough for the estimation. Thus, the improved performance of the SSVSP_DI could be a result of the reduced number of restrictions. However, only searching for CSH restrictions does not lead to improvements compared to the SSVSP. The SSVSP_DI captures the high probabilities for no dynamic interdependencies which are present in the data. The probability for homogeneity seems to be lower. Excluding dynamic interdependencies based on a specific threshold improves the forecast performance for the higher horizons. The two specifications, SSVSP_setDI_v1 and SSVSP_setDI_v2, can particularly well pick up the sparsity in the data. The model with a lower threshold value, SSVSP_setDI_v2, leads to higher improvements.

The performance of the SSVS is volatile, ranging from 9.52% to 95.24% of MSFEs below or equal to one. It performs well for the last three reported horizons. The SSVS also searches for dynamic interdependencies, thus, it is similar to the SSVSP_DI specification. With the exception that the SSVSP_DI distinguishes between domestic and foreign variables. The results of the SSVS support the finding that the MSFEs favor priors which capture the possibility of no dynamic interdependencies. The S^4 includes

Table 1.6: Posterior predictive density relative to unrestricted VAR model

	number of PLs ≥ 0 (in %)				
	horizon 1	horizon 2	horizon 4	horizon 6	horizon 12
SSVSP	33.33	28.57	47.62	23.81	38.10
SSVSP_DI	57.14	42.86	42.86	42.86	61.90
SSVSP_CSH	28.57	19.05	38.10	23.81	57.14
SSVSP_setDI_v1	28.57	28.57	47.62	47.62	42.86
SSVSP_setDI_v2	28.57	52.38	52.38	61.90	66.67
S^4	47.62	47.62	33.33	33.33	38.10
SSVS	47.62	38.10	57.14	71.43	61.90

Note: PL: posterior predictive density evaluated at the true observation y_{t+h}, compared to unrestricted VAR model. Forecasts for 12 horizons for Jan 2005 to end of the sample. Unrestricted VAR model: parameters drawn from unrestricted part of distributions. S^4: prior of Koop and Korobilis (2015b). SSVS: prior of George et al. (2008). SSVSP_DI: SSVSP with only DI restrictions. SSVSP_CSH: SSVSP with only CSH restrictions. SSVSP_setDI_v1: threshold 0.99 to set zero DI restrictions. SSVSP_setDI_v2: threshold 0.5 to set zero DI restrictions. Σ drawn from inverse Wishart distribution with T degrees of freedom and identity matrix plus sum of squared residuals as scaling matrix. 110,000 Gibbs draws, 10,000 disregarded as burn-in-phase.

DI but assumes a specific matrix structure which does not seem to be supported by the data.

Table 1.6 presents in percent the number of PL, in difference to the unrestricted VAR model, which are higher or equal zero. In general, a higher PL indicates a better performance since the posterior predictive density covers the true observation with a higher probability. The results are generally in line with the findings based on the MSFEs but differ in magnitude and also in horizons. In particular, the PL results favor the SSVSP_DI, SSVSP_setDI_v1, and SSVSP_setDI_v2 as well as the SSVS. In contrast to the extremely volatile MSFE results, the SSVS outperforms the SSVSP at all horizons. However, the SSVSP_DI exceeds the SSVS at two and is equally good at one out of five horizons. Again, the results point to the direction that no dynamic interdependencies are present in the data and a prior which can pick up these characteristics performs well.

Compared to the findings based on the MSFEs the prior specifications are less often able to outperform the unrestricted VAR model. This could be explained by a higher parameter uncertainty of the selection priors since they are a mixture of two distributions. The higher uncertainty is reflected in the posterior predictive density. The results are in general in line with Korobilis (2016) who also shows a high volatility in the performance as well as improved forecasting results for the SSVS compared to the S^4. However, combining the advantages of both priors, the panel dimension of the S^4 and the single restrictions search of the SSVS, in the SSVSP does not seem to pay

Table 1.7: 10 highest posterior probabilities for the restrictions

DI		SI		CSH		
α_{ij}^{lk}	$p(\alpha_{ij}^{lk}=0)$	ψ	$p(\psi_{ij}^{lk}=0)$	α_{jj}^{lk}	α_{ii}^{lk}	$p(\alpha_{jj}^{lk}=\alpha_{ii}^{lk})$
$\alpha_{UK,DE}^{IR,IP}$	1.00	$\psi_{JP,US}^{IP,IP}$	0.99	$\alpha_{JP,JP}^{IR,IP}$	$\alpha_{DE,DE}^{IR,IP}$	1.00
$\alpha_{JP,DE}^{IR,IP}$	1.00	$\psi_{JP,DE}^{IP,IR}$	0.99	$\alpha_{UK,UK}^{IR,IP}$	$\alpha_{JP,JP}^{IR,IP}$	1.00
$\alpha_{DE,JP}^{IR,IP}$	1.00	$\psi_{JP,US}^{IP,IR}$	0.99	$\alpha_{CA,CA}^{IR,IP}$	$\alpha_{JP,JP}^{IR,IP}$	1.00
$\alpha_{CA,JP}^{IR,IP}$	1.00	$\psi_{JP,DE}^{IP,CPI}$	0.99	$\alpha_{IT,IT}^{CPI,IP}$	$\alpha_{FR,FR}^{CPI,IP}$	1.00
$\alpha_{JP,US}^{IP,CPI}$	1.00	$\psi_{JP,DE}^{IP,CPI}$	0.99	$\alpha_{FR,FR}^{CPI,IP}$	$\alpha_{JP,JP}^{CPI,IP}$	1.00
$\alpha_{IT,DE}^{CPI,IP}$	1.00	$\psi_{JP,DE}^{IP,IP}$	0.99	$\alpha_{IT,IT}^{CPI,IP}$	$\alpha_{JP,JP}^{CPI,IP}$	1.00
$\alpha_{IT,FR}^{CPI,IP}$	1.00	$\psi_{DE,US}^{IP,IR}$	0.99	$\alpha_{UK,UK}^{IR,IP}$	$\alpha_{DE,DE}^{IR,IP}$	1.00
$\alpha_{UK,JP}^{CPI,IP}$	1.00	$\psi_{IT,DE}^{IP,CPI}$	0.99	$\alpha_{FR,FR}^{IR,IP}$	$\alpha_{JP,JP}^{IR,IP}$	1.00
$\alpha_{FR,JP}^{CPI,IP}$	1.00	$\psi_{DE,US}^{IP,IP}$	0.99	$\alpha_{CA,CA}^{IR,IP}$	$\alpha_{DE,DE}^{IR,IP}$	1.00
$\alpha_{US,JP}^{CPI,IP}$	1.00	$\psi_{DE,US}^{IP,CPI}$	0.99	$\alpha_{FR,FR}^{IR,IP}$	$\alpha_{DE,DE}^{IR,IP}$	1.00

Note: 10 highest posterior probabilities are presented for DI, SI, and CSH restrictions. The probabilities, $p(\alpha_{ij}^{lk}=0)$, $p(\psi_{ij}^{lk}=0)$, and $p(\alpha_{jj}^{lk}=\alpha_{ii}^{lk})$, are calculated as one minus the posterior means for $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w}. The probabilities measure the proportion of $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w} draws that equal zero averaged over all Gibbs sampler draws meaning that the coefficients are drawn from the restricted part of the distribution.

off due to the large number of restrictions to search for. As Korobilis (2016) shows the approach of Canova and Ciccarelli (2009) has no clear advantage, measured by forecasting performance, over the selection priors.

Tables 1.7 and 1.8 provide the ten highest and ten lowest posterior probabilities for the restrictions, $p(\alpha_{ij}^{lk}=0)$, $p(\psi_{ij}^{lk}=0)$, and $p(\alpha_{jj}^{lk}=\alpha_{ii}^{lk})$. The SSVSP can provide a detailed ranking on how likely a restriction should be set based on the data on a variable basis. The algorithm is able to detect a nuanced structure of the restrictions present in the data. Since the presented PVAR model serves as an illustration, the economic findings should not be over-interpreted.

The posterior probabilities provide evidence that restrictions are especially supported for the industrial production variable while it is vice versa for interest rates. Table 1.7 provides the ten highest posterior probabilities. Probabilities are high, indicated by ones, that no dynamic impacts of foreign lagged IP growth on interest rates and CPI growth exist. Additionally, industrial production growth seems to be fairly independent from other variables, shown by the high probabilities for no static interdependencies between IP growth and other variables. Finally, the probabilities for homogeneity of coefficients are especially high for the industrial production variables in other equations.

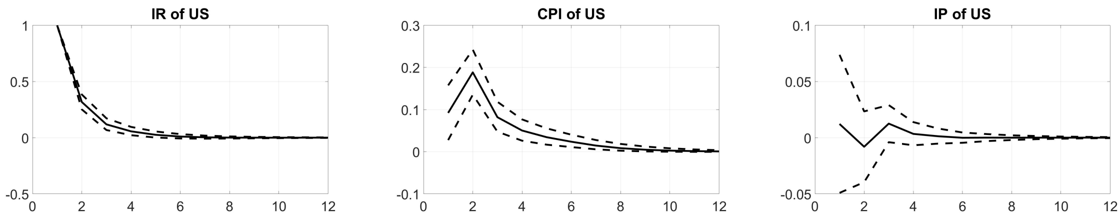
Table 1.8 presents the ten lowest posterior probabilities for restrictions. Lagged for-

Table 1.8: 10 lowest posterior probabilities for the restrictions

DI		SI		CSH		
α_{ij}^{lk}	$p(\alpha_{ij}^{lk}=0)$	ψ	$p(\psi_{ij}^{lk}=0)$	α_{jj}^{lk}	α_{ii}^{lk}	$p(\alpha_{jj}^{lk}=\alpha_{ii}^{lk})$
$\alpha_{FR,DE}^{IR,IR}$	0.00	$\psi_{CA,US}^{CPI,CPI}$	0.00	$\alpha_{DE,DE}^{CPI,CPI}$	$\alpha_{US,US}^{CPI,CPI}$	0.00
$\alpha_{CA,IT}^{IR,IR}$	0.00	$\psi_{UK,FR}^{CPI,CPI}$	0.00	$\alpha_{DE,DE}^{IP,IP}$	$\alpha_{US,US}^{IP,IP}$	0.06
$\alpha_{IT,DE}^{IR,IR}$	0.01	$\psi_{CA,FR}^{IR,IR}$	0.00	$\alpha_{FR,FR}^{IP,IR}$	$\alpha_{JP,JP}^{IP,IR}$	0.06
$\alpha_{DE,UK}^{IP,IR}$	0.01	$\psi_{UK,DE}^{IR,IR}$	0.00	$\alpha_{FR,FR}^{CPI,CPI}$	$\alpha_{US,US}^{CPI,CPI}$	0.07
$\alpha_{UK,US}^{IR,IR}$	0.03	$\psi_{DE,US}^{IR,IR}$	0.00	$\alpha_{IT,IT}^{IP,IR}$	$\alpha_{JP,JP}^{IP,IR}$	0.07
$\alpha_{FR,US}^{CPI,CPI}$	0.04	$\psi_{UK,US}^{IR,IR}$	0.00	$\alpha_{CA,CA}^{IP,IR}$	$\alpha_{JP,JP}^{CPI,CPI}$	0.07
$\alpha_{CA,US}^{IP,IR}$	0.04	$\psi_{CA,US}^{IR,IR}$	0.00	$\alpha_{UK,UK}^{IP,IR}$	$\alpha_{JP,JP}^{IP,IR}$	0.09
$\alpha_{FR,US}^{IP,IR}$	0.04	$\psi_{CA,FR}^{CPI,CPI}$	0.00	$\alpha_{JP,JP}^{IP,IR}$	$\alpha_{DE,DE}^{IP,IR}$	0.09
$\alpha_{FR,US}^{IP,IP}$	0.05	$\psi_{UK,JP}^{CPI,CPI}$	0.00	$\alpha_{JP,JP}^{IP,IR}$	$\alpha_{US,US}^{IP,IR}$	0.10
$\alpha_{CA,JP}^{IR,IR}$	0.05	$\psi_{FR,DE}^{CPI,CPI}$	0.00	$\alpha_{JP,JP}^{CPI,CPI}$	$\alpha_{DE,DE}^{CPI,CPI}$	0.10

Note: 10 lowest posterior probabilities are presented for DI, SI, and CSH restrictions. The probabilities, $p(\alpha_{ij}^{lk}=0)$, $p(\psi_{ij}^{lk}=0)$, and $p(\alpha_{jj}^{lk}=\alpha_{ii}^{lk})$, are calculated as one minus the posterior means for $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w}. The probabilities measure the proportion of $\gamma_{DI,ij}^{lk}, \gamma_{SI,ij}^{lk}$, and γ_{CSH}^{w} draws that equal zero averaged over all Gibbs sampler draws meaning that the coefficients are drawn from the restricted part of the distribution.

Figure 1.2: Responses of US variables to a shock to the US interest rate

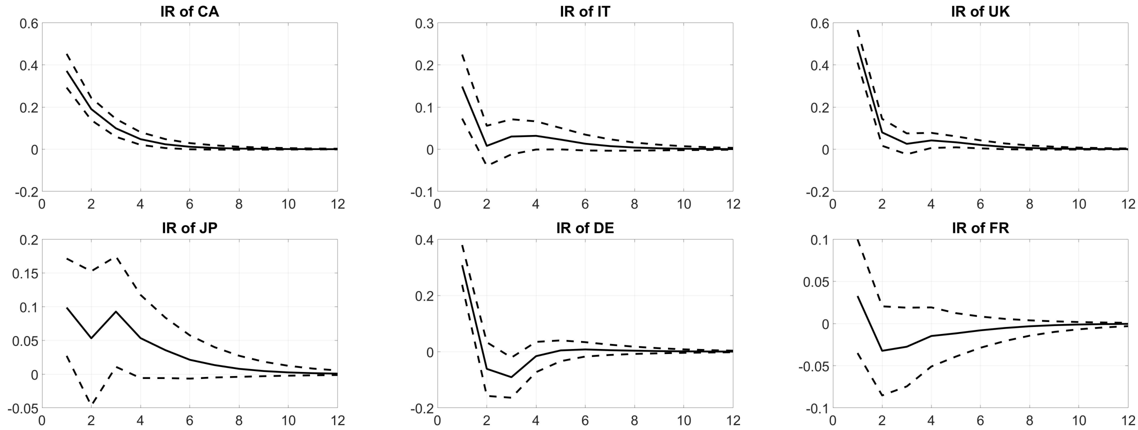


Note: Solid line shows response, dotted lines present 68% Bayesian credible interval.

eign interest rates seem to affect the domestic variables. Furthermore, US variables have a dynamic impact on other countries' variables. Both findings are supported by low probabilities for no DI. The lowest probabilities for the SI restrictions are found for combinations of the same variable, in particular for inflation and the interest rate. Heterogeneity is favored - low probability for no CSH - for the effect of inflation on inflation and of the interest rate on industrial production growth.

The impulse response analysis sheds light on the reliability of the findings. Exemplary, the responses to a shock to the US interest rate, presented in figure 1.2 for US variables and in figure 1.3 for foreign interest rates, is discussed. A contractionary US monetary policy, shown by an increase in the US interest rate, leads to a rise in US CPI growth in this system. The response of industrial production growth is insignificant. The increase of inflation in response to a tightening in the monetary policy is in line

Figure 1.3: Responses of foreign interest rates to a shock to the US interest rate



Note: Solid line shows response, dotted lines present 68% Bayesian credible interval.

with the price puzzle. The price puzzle refers to this result contradicting theoretical models and empirical findings which would claim that a rise in the interest rate leads to a decline in inflation. The puzzle is expected for VAR models which just include industrial production growth, inflation, and a short term interest rate and have a structural identification based on a recursive system. The finding of the price puzzle underlines that the estimated PVAR model here can only serve as an illustration and has its clear limitations.

The foreign interest rates immediately raise in response to a tightening in the US monetary policy. The increases in the interest rates are lower, below 0.5, than the initial raise in the US interest rate, which is normalized to one. The UK interest rate is initially affected most, followed by the Canadian and German interest rate responses. After around two horizons the effect of the US shock is insignificant for the interest rate of the United Kingdom, Germany, and Italy. The responses of the interest rates of Japan and France are lowest. For Japan the response is insignificant after the first horizon while for France the response is insignificant for all horizons. The raise in the Canadian interest rate lasts longest and comes to zero after propagation horizon six.

To sum up, the impulse response functions support that the results based on SSVSP are in line with theoretically expected responses from a recursively identified system with the three included variables. The illustrative model provides evidence that the results obtained using the SSVSP are plausible.

1.7 Conclusions

This paper introduces the SSVSP as an extension of the Bayesian S^4 proposed by Koop and Korobilis (2015b). The SSVSP is an alternative Bayesian estimation procedure for

PVAR models that is able to fully incorporate dynamic and static interdependencies as well as cross-country heterogeneities. It allows for a flexible panel structure since it only distinguishes between domestic and foreign variables. Using a hierarchical prior, the SSVSP searches for restrictions that are supported by the data.

The results of the Monte Carlo simulations demonstrate that the SSVSP outperforms the S^4 in terms of deviation from the true values in particular when a less restrictive panel structure is present. The average deviation of the estimated parameters from the true values for the simulation with a flexible panel structure is less for the SSVSP, $APD_{SSVSP} = 0.036$, than for S^4, $APD_{S^4} = 0.056$. The SSVSP_DI, where a mixture prior is only set on the parameters measuring dynamic interdependencies, has the smallest deviation from the true values of all models. Furthermore, the accuracy of the SSVSP in selecting the restrictions is proven by the posterior probabilities for no interdependencies and homogeneity.

The results of the empirical application are summarized in three main findings. First, the forecast performance is especially good for the SSVSP_DI and the two selection models, SSVSP_setDI_v1 and SSVSP_setDI_v2. Thus, restrictions search for no dynamic interdependencies is beneficial. However, the performance of the SSVSP is limited by the very large number of restrictions searched for. Second, posterior probabilities for DI and SI restrictions show that interest rates likely depend on foreign interest rates while foreign industrial production growth does not impact other domestic variables. Third, responses to a shock in the US interest rate are in line with expected impulse response functions.

The SSVSP prior can be further developed. The SI restrictions search, based on data, is an initial way to achieve structural identification, but it is limited by the fact it is built on a recursive system. Furthermore, in this specification, the hyperparameters are fixed for all parameters that are estimated. George et al. (2008) propose a default semi-automatic approach to select the hyperparameters which vary for each coefficient. Trying this approach leads to hyperparameters that tend to be so small that the majority of values are drawn from the loose part of the prior. Koop and Korobilis (2015b) specify distributions for the hyperparameters as also suggested in Giannone et al. (2015). This allows them to have varying and less subjectively chosen hyperparameters.

To sum up, the findings of the Monte Carlo simulations conducted and the exemplary empirical application encourage the use of the SSVSP to estimate PVAR models. However, further research regarding both the recursive structural identification and the specified hyperparameters should be undertaken.

Appendix

1.A Gibbs Sampler Algorithm

The full unrestricted PVAR model with one lag including N countries and for each country G variables can be written as

$$Y_t = Z_{t-1}\alpha + U_t,$$

where α is the vectorized $[NG \times NG]$-coefficient matrix A for lag one. The $Z_{t-1} = (I_{NG} \otimes Y_{t-1})$ where $Y_{t-1} = (y'_{1t-1}, ..., y'_{Nt-1})'$ and y_{it} denotes a vector of dimension $[G \times 1]$. The Y_t and U_t are $[NG \times 1]$-vectors. The U_t is normally distributed with mean zero and covariance matrix Σ that is of dimension $[NG \times NG]$. The element α_{ij}^{lk} refers to the coefficient of variable k of country j in the equation of variable l of country i.

The Gibbs sampler algorithm has the following three steps:

Step 1: Sample α from a normal posterior conditional on Σ, γ_{DI}, γ_{CSH}.

$$\alpha \mid \Sigma, \gamma_{DI}, \gamma_{CSH} \sim \mathcal{N}(\Gamma\mu_\alpha, V_\alpha),$$

where $V_\alpha = ((D'D)^{-1} + \Sigma^{-1} \otimes X'X)^{-1}$ with $X = Y_{t-1}$ and $\mu_\alpha = V_\alpha((\Sigma^{-1} \otimes X'X)\alpha_{OLS})$. D is a diagonal matrix with $D = diag(h_{11}^{11}, ..., h_{NN}^{GG})$. The value of h depends on γ_{DI} and γ_{CSH}:

$$h_{ij}^{lk} = \begin{cases} \tau_1, & if\ \gamma_{DI,ij}^{lk} = 0 \\ \tau_2, & if\ \gamma_{DI,ij}^{lk} = 1 \end{cases}$$

for the parameters, where DI restriction search is done $(i \neq j)$ and

$$h_{jj}^{lk} = \begin{cases} \xi_1, & if\ \gamma_{CSH}^{w} = 0 \\ \xi_2, & if\ \gamma_{CSH}^{w} = 1 \end{cases}$$

for the block diagonal parameters where CSH restriction search is done. α_{OLS} is the OLS estimate of α. The posterior mean is restricted with the selection matrix Γ.

Step 2: Update γ_{DI} and γ_{CSH} from Bernoulli distribution:

$$\begin{aligned}\gamma^{lk}_{DI,ij} &\sim Bernoulli(\pi^{lk}_{DI,ij})\\ \pi^{lk}_{DI,ij} &= \frac{u2^{lk}_{DI,ij}}{u1^{lk}_{DI,ij}+u2^{lk}_{DI,ij}}\\ \gamma^{w}_{CSH} &\sim Bernoulli(\pi^{w}_{CSH})\\ \pi^{w}_{CSH} &= \frac{v2^{w}_{CSH}}{v1^{w}_{CSH}+v2^{w}_{CSH}}.\end{aligned}$$

Hereby, $u1^{lk}_{DI,ij} = f(\alpha^{lk}_{ij} \mid 0, \tau_1^2) prob_{DI}$ and $u2^{lk}_{DI,ij} = f(\alpha^{lk}_{ij} \mid 0, \tau_2^2)(1 - prob_{DI})$. $f()$ denotes the probability density function of the normal distribution with mean zero and variance τ_1^2 or τ_2^2 evaluated at α^{lk}_{ij}. The parameter $prob_{DI}$ is set equal to 0.5. This shows that *a priori* the researcher assumes that it is equally likely that a dynamic interdependency between two variables of country i and j are zero or nonzero. $v1^{w}_{CSH} = f(\alpha^{lk}_{jj} \mid \alpha^{lk}_{ii}, \xi_1^2) prob_{CSH}$ and $v2^{w}_{CSH} = f(\alpha^{lk}_{jj} \mid 0, \xi_2^2)(1 - prob_{CSH})$. Again, $prob_{CSH}$ is set equal 0.5. Depending on γ^{w}_{CSH} the elements in Γ_w are updated.

Step 3: Update $\Sigma = \Psi^{-1'}\Psi^{-1}$ and γ_{SI}. The variance elements, ψ^{kk}_{ii}, are drawn from a Gamma distribution:

$$(\psi^{kk}_{ii})^2 \sim \mathcal{G}\left(a + \frac{T}{2}, B_n\right),$$

where $n = 1, ..., NG$ and

$$B_n = \begin{cases} b + 0.5SSE_{nn} & n = 1 \\ b + 0.5(SSE_{nn} - s_n'(S_{n-1} + (R'R)^{-1})^{-1}s_n) & n = 2, ..., NG \end{cases}.$$

Note that ψ^{11}_{11} is assigned to B_2, ψ^{22}_{11} to B_2, ..., and ψ^{GG}_{NN} to B_{NG}. T is defined as the length of the time series and SSE as the sum of squared residuals. S_n is the upper-left $n \times n$ submatrix of SSE, and $s_n = (s_{1n}, ..., s_{n-1,n})'$ contains the upper diagonal elements of SSE. R is a diagonal matrix with $R = diag(r^{11}_{11}, ..., r^{GG}_{NN})$. The value of r depends on γ_{SI}:

$$r^{lk}_{ij} = \begin{cases} \kappa_1, & if\ \gamma^{lk}_{SI,ij} = 0 \\ \kappa_2, & if\ \gamma^{lk}_{SI,ij} = 1 \end{cases}.$$

Define the vector $\psi = (\psi^{11}_{12}, ..., \psi^{GG}_{N-1,N})'$. Thus, ψ contains the covariance elements, ψ^{lk}_{ij} for all $i \neq j$ and has the dimension $[n_{SI} \times 1]$, where $n_{SI} = 1, .., N_{SI}$ and N_{SI} is the length equal to the number of SI restrictions. The elements of ψ are updated from a normal distribution:

$$\psi_{n_{SI}} \mid \alpha, \psi, \gamma_{SI} \sim \mathcal{N}(\mu_{n_{SI}}, V_{n_{SI}}).$$

Hereby, $\mu_{n_{SI}} = -\psi_{ii}^{kk}(S_{n_{SI}-1} + (R'R)^{-1})^{-1}s_{n_{SI}}$ and $V_{n_{SI}} = (S_{n_{SI}-1} + (R'R)^{-1})^{-1}$. The element ψ_{ii}^{kk} is the variance element in the same row of Ψ as $\psi_{ij}^{lk} = \psi_{n_{SI}}$ for all $i \neq j$. The off-diagonal elements of the covariance matrix that belong to one country are drawn from a normal distribution with mean zero and variance κ_2.

1.B Hyperparameter

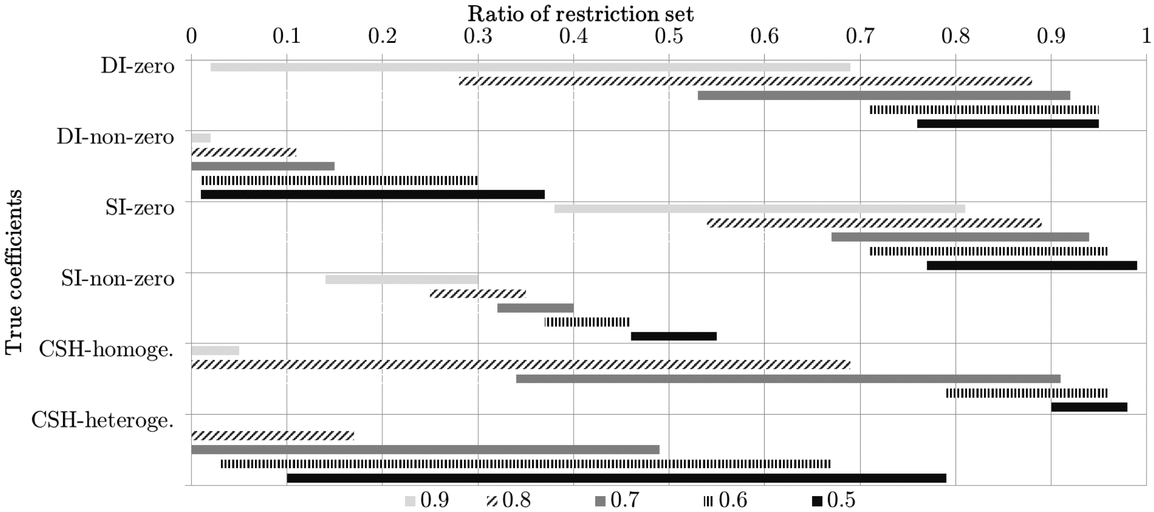
Table 1.9: Hyperparameters

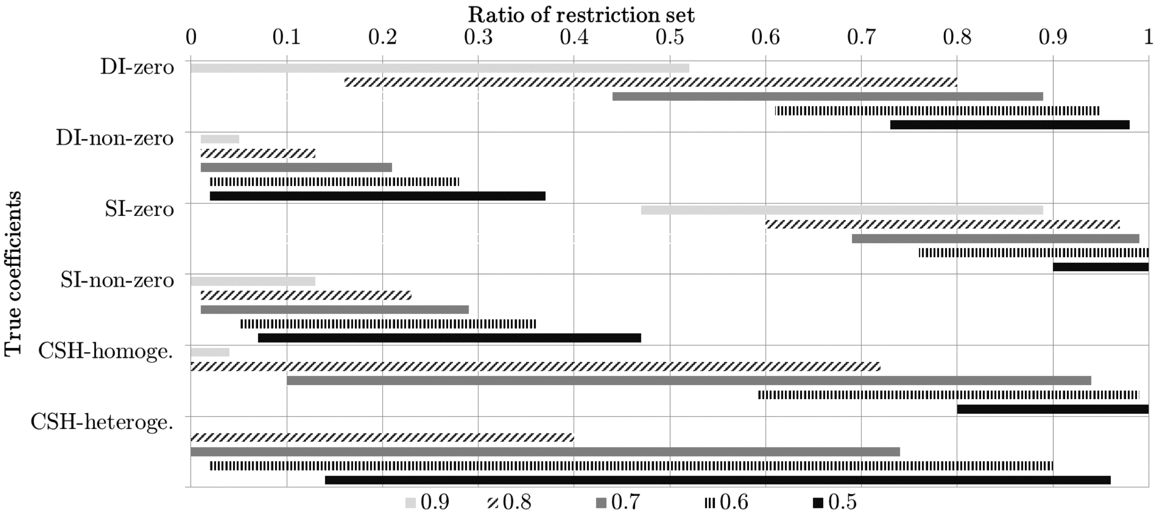
τ_1	τ_2	ξ_1	ξ_2	κ_1	κ_2	a	b
0.2	4	0.2	4	0.3	4	0.01	0.01

A value of $\tau_1 = 0.2$ and $\tau_2 = 4$ means that the variance of the tight prior equals 0.04 and 16 for the loose prior. The criterion that the variance of the first part of the normal distribution is smaller than the second part is clearly fulfilled. Several other specifications are also checked. The accuracy of the algorithm in selecting the restrictions varies with the specification of the hyperparameters. If the τ_1, κ_1, and ξ_1 are chosen too small, the majority of values is drawn from the second part of the normal distribution (γ equals one with a very high probability). Still, γ equals more often one in the cases no restriction is set in the true specification of the Monte Carlo simulation. Values for hyperparameters smaller than or equal to 0.1 prove to be too small, resulting in the difficulties mentioned. George et al. (2008) propose a default semi-automatic approach to selecting the hyperparameters. The values are not fixed, but varying for each coefficient. For example $\tau_{1,i} = c_1\sqrt{var(\alpha_i)}$ and $\tau_{2,i} = c_2\sqrt{var(\alpha_i)}$ whereby $c_1 = 0.1$ and $c_2 = 10$. $var(\alpha_i)$ is the OLS estimate of the variance of the coefficient in an unrestricted model. κ and ξ are set in an equal manner. Trying this approach also leads to hyperparameters smaller than 0.1. The hyperparameters of the other priors are set to the proposed default values of the authors. For S^4 the small variance values are set to 0.1 and the high variance values to square root of 10, for the SSVS small variance values are set to 0.1 and high variance values to 5 as used by Koop and Korobilis (2015b) and George et al. (2008).

1.C Monte Carlo Simulation

Figure 1.4: Share of restrictions set according to threshold value - simulations 1 and 2





Note: First graph shows results for simulation one, second for simulation two. Number of draws of $\gamma^{lk}_{DI,ij}, \gamma^{lk}_{SI,ij}$, and γ^{w}_{CSH} that equal zero averaged over all Gibbs sampler draws as a ratio of all simulated samples for a given threshold. True coefficients are the parameters set when simulating the DGPs. Simulation 1: DGP has matrix panel structure. Simulation 2: DGP has flexible panel structure.

CHAPTER 2

Penalized Estimation of Panel Vector Autoregressive Models: A Lasso Approach

2.1 Introduction

Growing international interlinkages in the financial and real sector are a defining feature of the global economy and have risen in importance over recent decades. This involves major economic policy implications as highlighted for example by numerous IMF reports and notes on spillovers. Theoretical papers demonstrate that ignoring international spillovers could lead to biased impulse response functions and to inaccurate forecasts. Georgiadis (2017) stresses that the accuracy of spillover effects increases significantly when they are estimated with multi-country models instead of bilateral vector autoregressive models. Furthermore, leaving out variables capturing international connections could lead to an omitted variable bias impacting structural analyses, as discussed, for example, by Kilian and Lütkepohl (2017). In addition, Pesaran et al. (2009) point out that not accounting for linkages across countries can lead to less accurate forecasts of macroeconomic variables. Consequently, multi-country models with several variables, such as panel vector autoregressive (PVAR) models, are necessary to capture global spillovers in economic analyses.

The strength of PVAR models is to account for interdependencies and heterogeneities across nations by jointly modeling multiple variables of several economies. PVAR models enable the modeling of dynamic interdependencies by augmenting country specific models with lagged foreign variables. These models allow for static interdependencies measured by potential nonzero covariances between the error terms of different countries. Moreover, PVAR models take cross-country heterogeneities into account by specifying country-specific coefficient matrices. However, estimating these models is challenging because a large number of parameters is usually set against a short time series. Due to the curse of dimensionality, estimation of these models is thus often infeasible.

This paper, first, proposes a new lasso (least absolute shrinkage and selection operator) that is suitable for estimating PVAR models and, second, establishes the asymptotic oracle properties of this estimator. The proposed estimator modifies the standard lasso in two respects. First, the lasso for PVAR models takes the panel structure inherent to the data into account. This is achieved by introducing penalty terms incorporating time series and cross-sectional properties. Second, it allows for an unrestricted covariance matrix at the same time. This is done by specifying the loss function of the estimation problem as the weighted sum of squared residuals, thereby, accounting for the correlation between error terms of different cross section units. Thus, the lasso provides a frequentist solution to the curse of dimensionality problem by using the panel structure to ensure the estimation feasibility.

Furthermore, this paper establishes the asymptotic oracle property of the lasso for PVAR models. That means, asymptotically the lasso selects the true variables to be included in the model and estimates nonzero parameters as efficiently as if the true underlying model is known. This property is an important feature for a variable selection estimator and is not achieved for the standard lasso. Since the newly introduced penalty term varies across variables and is therefore closely related to the adaptive lasso, the asymptotic oracle property can be derived.

In general, the lasso, as proposed by Tibshirani (1996), regulates the dimension of the model by constraining the estimation problem with a linear penalty term. The penalization determines the sum of the absolute values of the regression coefficients, that is, the L_1-norm of the coefficient matrix, to be less than a fixed value. Thus, the penalty term governs the degree of shrinkage. By forcing some coefficients to be zero and shrinking others, the lasso chooses the variables to be included in the model.

The main advantages of the lasso technique applied here are threefold.[1] First, the specified penalty parameters of the lasso for PVAR models account for the inherent panel structure within the data. The penalty terms build on a specific expected structure in panel data models. That is, interdependencies are assumed to only exist between specific variables and cross section combinations and decrease over lags. This structure is in line with the specification of the Minnesota prior for vector autoregressive (VAR) models. The lasso uses this structure to reduce the dimension of the parameter space by setting specific coefficients to zero. In particular, the penalty terms capture that more recent lags provide more important parts of the dynamics than more distant ones and that lags of domestic variables are more important than lags of foreign variables.

[1]Other methods to ensure the feasibility of the estimation are factor approaches, Bayesian shrinkage priors, selection priors, and classical shrinkage methods, such as the ridge regression. Some methods that have been used for PVAR models are described in section 2.5 in detail.

As demonstrated by Song and Bickel (2011) and Nicholson et al. (2016, 2017), including grouping structures or time series properties in the specification of the lasso for estimating VAR models can improve forecast accuracy compared to the normal lasso penalty. The authors let the penalty term vary across lags and include grouping structures by using group lasso techniques as proposed by Yuan and Lin (2006). This allows them to capture similar sparsity patterns in the coefficient matrix. Likewise, contributions on Bayesian selection priors for PVAR models support that accounting for the inherent panel dimension within the data can enhance forecasting performance.[2]

Second, considering an unrestricted covariance matrix by the specification of the loss function includes possible correlations between error terms in the estimation of the parameters. In penalized regressions, coefficients derived using generalized least squares deviate from those derived by ordinary least squares. Using the sum of squared residuals as the loss function disregards possible correlations between variables, thereby restricting the covariance matrix to the identity matrix. Hence, this procedure imposes strict assumptions on the dependence structure between the cross-sectional units. Lee and Liu (2012) show this for the use of lasso for VAR models. Basu and Michailidis (2015), Ngueyep and Serban (2015), and Davis et al. (2016) modify the loss functions in the lasso optimization for VAR models and allow for unrestricted covariances in the penalized estimation.

Third, the lasso for PVAR models benefits from the same properties as the lasso proposed by Tibshirani (1996). That is, the lasso reduces the dimension of the estimated model. Thereby, it ensures the feasibility of the estimation if the number of parameters per equation exceeds the number of observations. Furthermore, the lasso simultaneously selects and estimates the model. It allows for a flexible lag structure across equations since the lasso can choose different lag orders for each equation of the model. Moreover, the lasso is able to improve forecast prediction accuracy by reducing the variance of the predicted values.

The lasso for PVAR models is of interest for empirical work since it provides a solution to ensure the estimation feasibility for PVAR models. That is relevant since, first, PVAR models are typically large including several countries and variables per country to capture macroeconomic relations. Second, the dimension of PVAR models grows fast as adding a country increases the number of equations and columns of the coefficient matrices while adding variables means including them for each country. The lasso for PVAR models can be used for estimating reduced form VAR models. It can select the subset of variables that should be included in the model and serve as a flexible lag length selection tool. Due to the selection of the relevant variables the

[2] See, for example, Koop and Korobilis (2015b), Korobilis (2016) and Schnücker (2016).

PVAR model estimated via lasso is easily interpretable and might be used for further structural analysis or forecasting.

The results of three simulations and an empirical application support the use of the lasso for PVAR models. It improves the forecast accuracy measured by mean squared forecast errors relative to estimating the PVAR model with OLS, relative to Bayesian panel VAR methods, and relative to single country models. Accounting for the panel dimension in the penalty terms increases the forecast performance as using a lasso approach without such specific penalty terms leads to larger mean squared forecast errors. The gain in forecast accuracy relative to other estimation techniques is, in particular, found for large systems in simulations and an empirical application. For smaller models, the lasso for PVAR models performs equally to the models of comparison.

Furthermore, the dimension reduction of the lasso techniques results in smaller mean squared errors, measuring the deviation of coefficient estimates from their true values, for all simulations compared to OLS. The benefit in terms of lower mean squared error is higher for large and sparse models. The lasso provides a frequentist alternative to estimate PVAR models which is competitive to alternative techniques. This is supported by mean squared errors of the lasso techniques in the same range with Bayesian panel VAR methods and single country models.

In the following, the paper is related to the relevant literature in section 2.2. In section 2.3 the lasso for PVAR models is introduced and its asymptotic properties are discussed. Other estimation strategies for PVAR models are reviewed. Next, in section 2.4 three simulation studies evaluate the performance of the lasso for PVAR models along different criteria. A forecasting exercise on example data is conducted in section 2.5 while section 2.6 concludes.

2.2 Literature

By introducing the lasso for PVAR models, this paper contributes, first, to the literature on the use of the lasso techniques for VAR models and, second, to the literature on estimation procedures for PVAR models.

Hsu et al. (2008) establish the usage of the lasso for VAR models. The authors, along with Kascha and Trenkler (2015), report that the lasso improves forecast performance compared to the use of information criteria for model selection. Ren and Zhang (2010) and Ren et al. (2013) build on Zou (2006), who propose adaptive weights for penalizing coefficients differently, and develop adaptive lasso techniques for VAR models. Ren and Zhang (2010) propose the adaptive lasso and the hybrid adaptive lasso which first determines the lag order by some information criterion for VAR models. Ren et al. (2013) build on the adaptive lasso and propose a two step adaptive lasso leading to unbiased estimates of the nonzero coefficients. Their results provide evidence that the adaptive lasso outperforms the lasso in terms of forecasting performance, thus indicating the benefit of coefficient specific penalties. Kock and Callot (2015) establish non-asymptotic oracle inequalities for the lasso and adaptive lasso for high-dimensional VAR models. The authors further show that the lasso provides asymptotically consistent estimates and that the adaptive lasso is asymptotically equivalent to the least squares estimator that only includes true nonzero parameters.[3]

To date, two main extensions of Tibshirani's lasso are proposed in the context of VAR models. As mentioned, one strand of the literature broadens the specification of the penalty term to include special characteristics. The second group modifies the loss function in order to allow for unrestricted covariance matrices. However, the papers are either part of the first or the second group. One exception is Ngueyep and Serban (2015), who propose a penalized log-likelihood scheme applying penalties for higher lags and within group or between group penalties. Thus, the authors take into account the covariance matrix and allow for special characteristics. Yet, they still restrict the covariance matrix in their approach to a block structure by assuming no dependence across groups. This paper fills the gap by combining the weighted sum of squared residuals as the loss function with penalty terms that incorporate data properties. In addition, the paper broadens the specification of the penalty terms used so far by the introduction of the new penalty term for PVAR models.

The first group consists of papers which further develop the penalty term. Song and Bickel (2011) and Nicholson et al. (2016, 2017) estimate VAR models with the lasso but use penalty parameters which are able to incorporate time series properties or grouping structures in the coefficient matrix. They do so by letting the penalty term vary across lags. The authors include a grouping structure by using group lasso techniques as proposed by Yuan and Lin (2006). This allows them to capture similar sparsity patterns in the coefficient matrix. Song and Bickel (2011) present different grouping structures: an universal grouping which sets lasso penalties on the diagonal and group lasso penalties on the off-diagonal elements and segmented grouping combining lasso, group lasso and group lasso penalties for subgroups. The variants additionally penalize increasing lag length. Nicholson et al. (2016) propose a hierarchical VAR model as an alternative to lag length selection methods. The authors describe different penalization structures for

[3]This paper focuses on the lasso estimated in a frequentist way and does not cover Bayesian lasso approaches. Bayesian lasso variants are, for example, discussed by Park and Casella (2008) and Kyung et al. (2010). Korobilis (2013), Gefang (2014), and Billio et al. (2016) use Bayesian lasso approaches for VAR models. Additionally, papers use the lasso for panel data regressions. Since this paper concentrates on the estimation of panel VAR models, these approaches are not further discussed. Other contributions include Ando and Bai (2016) and Su et al. (2016).

lags: the lag length can vary across equations but not within equations and one variant with milder assumptions on diagonal elements. Nicholson et al. (2017) demonstrate the forecast performance of various penalty schemes for VAR models with exogenous variables. The penalty schemes are a group lasso for each lag of endogenous variables and one for exogenous variables as well as a sparse group lasso allowing within group sparsity. The three studies compare the different penalty structures to models using the normal lasso approach. The findings of all three studies demonstrate that including time series and grouping characteristics lead to improved forecast performance.

Papers in the second group modify the loss function and generate weights for the sum of squared residuals. Lee and Liu (2012) explain that the choice of the loss function (sum of squared residuals or weighted sum of squared residuals) is crucial in the context of VAR models. In a VAR model the covariance matrix impacts the estimated parameters in a constraint regression. Using the sum of squared residuals as the loss function disregards possible correlation between variables and thereby restricts the covariance matrix to the identity matrix. Lee and Liu (2012), Basu and Michailidis (2015), Ngueyep and Serban (2015), and Davis et al. (2016) use a weighted sum of squared residuals as their loss function and hence allow for an unrestricted covariance matrix. As yet, the literature on estimating VAR models with the lasso follows two main approaches to estimate the covariance matrix: a two-step approach or a joint likelihood approach. Lee and Liu (2012) describe two plug-in methods, where in a first step either the coefficient matrix or the covariance is estimated, followed by the estimation of the other. The authors use a graphical lasso (glasso), following, in particular, Friedman et al. (2008). In addition, they present a doubly penalized likelihood approach to jointly estimate the coefficient and covariance matrix in a L1-regularized regression. Basu and Michailidis (2015) propose another option by estimating the covariance matrix using residuals of an initial lasso estimation with sum of squared residuals or a glasso approach. Further, they present a joint penalized maximum likelihood approach. Davis et al. (2016) compare their two-stage approach using tools from the frequency domain with a lasso approach weighted with the inverse covariance matrix. Updating until convergence, the covariance matrix is estimated using the residuals of the lasso estimation. Ngueyep and Serban (2015) propose a penalized log-likelihood scheme applying penalties for higher lags and within group or between group penalties.

Furthermore, the paper extends the current literature on the estimation of PVAR models. As yet, the literature mainly uses three kinds of model selection methods: the cross-sectional approach of Canova and Ciccarelli (2004, 2009), the Bayesian selection prior of Koop and Korobilis (2015b), and a priori assumptions of no dependence or homogeneity across the panel units. A detailed description of the alternative estimation strategies is provided in 1.2 of chapter 1 of this thesis. Some alternative estimators for PVAR models are described in 2.3.5.

2.3 The lasso for PVAR Models

The lasso for PVAR models modifies the traditional lasso of Tibshirani (1996) in two ways. First, a specific penalty term is introduced which captures time series and cross-sectional properties. Second, the loss function of the lasso optimization problem is weighted with the inverse covariance matrix, thereby allowing for an unrestricted covariance matrix.

2.3.1 PVAR Model

Panel vector autoregressive models include several countries and country-specific variables in one model. A PVAR model with N countries and G variables per country for $t = 1, ..., T$ is given by

$$y_{it} = A_{i1}Y_{t-1} + A_{i2}Y_{t-2} + ... + A_{ip}Y_{t-p} + u_{it}, \tag{2.1}$$

where y_{it} denotes a vector of dimension $[G \times 1]$ for country i with $i = 1, ..., N$.[4] The $Y_{t-P} = (y'_{1t-P}, ..., y'_{Nt-P})'$ are of dimension $[NG \times 1]$ and the coefficient matrices A_{iP} of dimension $[G \times NG]$ for $P = 1, ..., p$. The $u_{it} \sim \mathcal{N}(0, \Sigma_{ii})$ and the covariance matrices across countries are given by Σ_{ij} for $i \neq j$.

In compact form, the PVAR model can be written as

$$Y_t = BX_{t-1} + U_t, \tag{2.2}$$

where $Y_t = (y'_{1t}, ..., y'_{Nt})'$ and the coefficient matrix B is of dimension $[NG \times NGp]$. The vector X_{t-1} includes all lagged variables, $X_{t-1} = (Y_{t-1}, ..., Y_{t-p})'$, and is of dimension $[NGp \times 1]$. The U_t is normally distributed with mean zero and covariance matrix Σ of dimension $[NG \times NG]$. The unrestricted PVAR model allows for dynamic and static interdependencies as well as for heterogeneities across countries. The X_{t-1} includes lagged values of every variable in each equation. The unrestricted B-matrix and the covariance matrix Σ enable country-specific coefficients and correlations between error terms of all possible variable-country combinations. This PVAR model has $(NG)^2p$ unknown parameters of the B-matrix and $\frac{NG(NG+1)}{2}$ parameters of Σ. Variables are ordered per country meaning that the first G rows of the system model variables of

[4] Although this specification does not include a constant as common in the lasso literature since data are usually standardized, it can be extended to include one.

country one, while the rows $NG - G + 1$ to NG describe the variables of country N. The large number of parameters can lead to the curse of dimensionality problem. The lasso provides a solution to deal with this issue.

2.3.2 The lasso Estimator

Tibshirani (1996) proposed the lasso for a linear regression model with multiple regressors. The coefficient estimates are obtained by minimizing the sum of squared residuals subject to a linear constraint. The penalization term regulates the sum of the absolute values of the regression coefficients, the L_1-norm of the coefficients, to be less than a fixed value. The lasso forces the coefficients to lie in a specific area that is centered around zero. Thereby, it shrinks some coefficient and constrains others to be equal to zero. The L_1-norm determines the geometric shape of this constrained region. It has two properties that are crucial for the features of the lasso. Coefficients can equal zero due to the possibility of corner solutions and, second, the constrained region is convex, which simplifies the optimization procedure.

Introducing a shrinkage penalty in the regression enables coping with situations in which $T < NGp$, can improve prediction accuracy, and produce interpretable models.[5] If $T < NGp$, the number of parameters per equation exceeds the number of observations, ordinary least squares is not feasible since no unique solution exists. If the true model is sparse, meaning that some of the true coefficients are zero, the lasso finds a solution by constraining the estimation. Furthermore, the lasso reduces the variance of the estimated coefficients, thereby improving prediction accuracy. Due to the selection property of the lasso the interpretation of the model is enhanced. By setting some coefficients to zero, a subset of variables that simplifies the identification of core driving variables of the system is selected.

The three mentioned properties clarify for which situations the lasso is well suited, namely for large, sparse systems for which the researcher's aim is to provide forecasts and to analyze main driving forces. The bias introduced by the lasso is accepted in order to gain these properties.

2.3.3 Extended Penalty Term and Loss Function for PVAR Models

The optimization problem of the lasso for PVAR models modifies the lasso of Tibshirani (1996) in two ways. The weighted sum of squared residuals is used as the loss function instead of the sum of squared residuals. Furthermore, a penalty term capturing the time series and cross section properties is introduced. The resulting optimization problem

[5] Tibshirani (1996) and Hastie et al. (2015) discuss these three properties in detail.

is given by:

$$\begin{aligned} \underset{b_{km}}{\operatorname{argmin}} \quad & \frac{1}{T}\sum_{k=1}^{K}\sum_{j=1}^{K}\omega_{kj}\left(Y_k - \sum_{m=1}^{Kp} b_{km}X_m\right)\left(Y_j - \sum_{m=1}^{Kp} b_{jm}X_m\right)' \\ & + \sum_{k=1}^{K}\sum_{m=1}^{Kp}\lambda_{km}\left|b_{km}\right|, \end{aligned} \tag{2.3}$$

where b_{km} is the element of the B-matrix in the k-th row and m-th column. K is the number of countries times the number of variables for each country, $K = NG$. The Y_j and X_m are of dimension $[1 \times T]$ for $j = 1, ..., K$ and $m = 1, ..., Kp$. The ω_{kj} is an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$. The λ_{km} is the penalty parameter and $|b_{km}|$ denotes the absolute value of b_{km}. As common in the lasso literature, Y_t is demeaned and standardized. The latter is done in order to have comparable units for all variables when choosing the penalty parameters.

Furthermore, the stability of the VAR model is assumed. This assumption is needed for the derivations of the asymptotic results (see 2.3.4). To model data with cointegration relations a vector error correction model is more suitable. However, the estimation of the cointegration relations is not straightforward in the case of the proposed lasso. Here, an extension of the estimation procedure along the lines of for example Liao and Phillips (2015) would be necessary. Liao and Phillips (2015) propose a lasso for vector error corrections models and apply a penalty on the coefficient matrix and a penalty on the cointegration rank. The author set a group lasso penalty on the coefficient matrix of the lagged differences. Furthermore, they use an adaptive penalty function for rank selection penalizing a function of the eigenvalues of the cointegration matrix.

The optimization problem of the lasso for PVAR models is solved using a coordinate descent algorithm as proposed in Friedman et al. (2007) and Friedman et al. (2010).[6] This iterative algorithm updates from B_{iter}, the coefficient matrix B in iteration $iter$, to B_{iter+1} by a univariate minimization over a single b_{km}. It iterates over all elements in B till convergence is reached.[7] The coordinate descent algorithm can be used since the non differentiable part of the optimization problem is separable. Convexity and separability of the problem ensure the existence of a global solution. The lasso estimator,

[6]The optimization algorithm and the derivation of the lasso estimator are described in detail in 2.C and 2.A. For more details regarding the optimization algorithm see Friedman et al. (2007), Friedman et al. (2010) and Hastie et al. (2015).

[7]Convergence is achieved when $max(|B_{iter} - B_{iter-1}|) < \epsilon$. The ϵ is chosen such that the lasso solution converges to the OLS estimate for a penalty parameter set to zero and weighted sum of squared residuals as the loss function.

which is called *lassoPVAR* in the following, has the form:

$$b_{km}^{lasso} = sign\left(\tilde{b}_{km}\right)\left(\left|\tilde{b}_{km}\right| - \frac{\lambda_{km}T}{2\omega_{kk}X_mX_m'}\right) \tag{2.4}$$

with

$$\tilde{b}_{km} = \frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j - \sum_{i=1}^{Kp} b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k - \sum_{i\neq m}^{Kp} b_{ki}X_i)X_m'}{X_mX_m'}. \tag{2.5}$$

As pointed out by Lee and Liu (2012), in a VAR model correlations between error terms have an impact on the estimated parameters in a restricted regression.[8] It can be easily seen from the above stated lasso estimator b_{km}^{lasso} that the covariance affects the value of b_{km}^{lasso} for elements $\omega_{kk} \neq 1$ and $\omega_{jk} \neq 0$ for $j \neq k$. When Σ equals the identity matrix, the estimator b_{km}^{lasso} reduces to the lasso estimator based on the sum of squared residuals as the loss function.

The covariance matrix Σ is estimated using a two-step approach. The first step estimates the covariance matrix via graphical lasso, while in the second step the estimated $\hat{\Sigma}$ is plugged into the lasso estimation of b_{km}^{lasso}. Friedman et al. (2008) demonstrate that the covariance matrix is estimated by maximizing the Gaussian penalized log-likelihood

$$\log\det(\Omega) - tr(S\Omega) - \rho\,||\Omega|| \tag{2.6}$$

with respect to the nonnegative definite inverse of the covariance matrix $\Omega = \Sigma^{-1}$. The matrix S is the empirical covariance, $tr(S\Omega)$ is the trace of $S\Omega$ and $||\Omega||$ is the sum of the absolute values of each element of Ω. For $\rho > 0$ the regression is penalized, while for $\rho = 0$ the classical maximum likelihood estimator is obtained. The details of the glasso are in 2.B. As pointed out by Banerjee et al. (2008), $\hat{\Sigma}$ is invertible even in the case when the number of variables is larger than the number of observations due to the regularization using $\rho > 0$.

An alternative way to estimate the covariance matrix, as done by, for example, Tibshirani (1996), is to use the least squares estimator

$$\hat{\Sigma} = \frac{1}{T - kk}(Y - \hat{B}X)(Y - \hat{B}X)',$$

where kk is the number of degrees of freedom. The degrees of freedom adjusted least squares estimator is a consistent estimator for constrained regression problems, although zero restrictions can reduce the number of degrees of freedom. Another option is to use the number of degrees of freedom for the lasso, which is the number of nonzero

[8]See Lee and Liu (2012) for details. This is similar to the well-known fact that for VAR models, OLS is unequal to GLS in the case of parameter constraints.

parameters.[9] However, in contrast to the glasso estimation, this approach can lead to problems for the invertibility of the covariance matrix in large systems. This is why the glasso approach is used here.

The second extension of the lasso for PVAR models is the modification of the penalty term. The λ_{km} denotes the penalty parameter. If $\lambda_{km} = 0$, the estimated coefficients equal the OLS solutions. If $\lambda_{km} > 0$, the parameters are shrunk to zero. To allow for a specific time series and cross section penalty, λ_{km} consists of three parts:

$$\lambda_{km} = \lambda_k \; p^{\alpha} \; c. \tag{2.7}$$

1. **Basic penalty** - λ_k. This part varies across equations. $\lambda_k > 0$ will force coefficients to zero.
2. **Time series penalty** - p^{α}. It captures that more recent lags provide more information than more distant ones. The penalty increases with the lag order, p, for $\alpha > 0$. The time series penalty part allows the penalty to vary across lagged variables.
3. **Cross section penalty** - $c > 1$, if foreign variable. The penalty models that lags of domestic variables have a larger impact than lags of foreign variables.

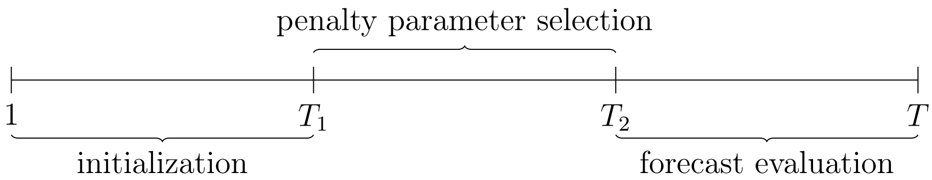
The penalty parameters vary across equations (due to λ_k) and across lagged variables (due to p^{α} and c). The parameters α and c are fixed for the whole model.

The cross section penalty separates domestic and foreign variables. However, all foreign variables are treated in a similar way as c is fixed for the whole model. That is done to simplify the selection of the penalty parameter. A c varying across different countries complicates the determination of the optimal penalty parameters and increases the computational time. For some empirical application a more flexible c might be appropriate. Such a flexibility can be build in by grouping countries and having different c parameters for sub-groups of countries. Following the idea of GVAR models, the c parameter can also be modeled depending on exogenous connectivity measures such as trade weights.

Optimal penalty parameters are determined via a rolling cross-validation technique. The penalty parameters are chosen such that they minimize one-step ahead mean squared forecast errors.[10]

[9] Regarding the degrees of freedom for the lasso see Bühlmann and van de Geer (2011) for details.

[10] The n-fold cross-validation technique for choosing the optimal penalty parameter is not applied here due to the time dependence in the PVAR model. By choosing the optimal penalty parameter that minimizes one-step ahead mean squared forecast errors, this paper follows Song and Bickel (2011), Nicholson et al. (2016, 2017). However, in contrast, Bergmeir et al. (2018) justify the use of the standard n-fold cross-validation techniques for autoregressive processes.



Like Song and Bickel (2011), the sample is split in three periods: The first period from 1 to $T_1 - 1$ is used for estimating the model, based on the second period from T_1 to $T_2 - 1$ different penalty parameters are evaluated, and the third period from T_2 to the end of the sample is later used for forecast evaluation of the lasso.[11] The model is estimated in a rolling scheme taking the observations from t to $T_1 + t - 1$ for $t = 1, ..., T_2 - T_1$. For each t the out-of-sample forecast accuracy for a specific penalty parameter λ_{km} is measured by the one-step ahead mean squared forecast error for variable k, $k = 1, ..., NG$:

$$MSFE(\lambda_{km})_k = \frac{1}{T_2 - T_1} \sum_{t=1}^{T_2 - T_1} (\hat{Y}_{k,t+1} - Y_{k,t+1})^2,$$

where $\hat{Y}_{k,t+1}$ denotes the estimated one-step ahead forecast for variable k. For simplicity only λ_k is determined via cross-validation, while α and c are pre-set to $\alpha = 0.6$ and $c = 1.4$ for the simulation and $\alpha = 0.6$ and $c = 1.8$ for the application. These values are preselected in a small cross-validation exercise. The search for the optimal λ_k is done over a grid of penalty parameter values whereby at the maximal value all coefficients equal zero.[12] The forecast performance is evaluated for the period T_2 to T by MSFEs based on rolling window forecasts with the fixed penalty parameters determined for the period 1 to $T_2 - 1$.

The application of the lasso for PVAR models is not limited to the currently considered PVAR model in which the cross sections are countries. Other possible cross-sectional dimensions are, for example, industries and regions. More generally, the cross section penalty can be understood as a higher penalty for variables of a cross section unit different than the one of the variable being explained.

[11] For estimation, for the simulation the periods are $T_2 = T - 20$ and $T_1 = T_2 - 20$ and for the application $T_2 = T - 20$ and $T_1 = T_2 - 60$. Extending the period for penalty parameter selection comes at the cost of longer computational time.

[12] For the simulations: $\lambda_k^{max} = max(max(XY'))$ and λ_k^{grid} are six values between 0.01 and $(1/NGp)\lambda_k^{max}$. For the applications: $\lambda_k^{max} = max(max(XY'))$ and λ_k^{grid} are twelve values between 0.01 and $(1/T)\lambda_k^{max}$. See 2.F.1 for details on the grid values for the application.

2.3.4 Asymptotic Properties

As a variable selection method, the lasso for PVAR models should satisfy the oracle properties.[13] This means, first, asymptotically the lasso selects the correct sparsity pattern. With probability tending to one, it sets true zero parameters to zero while not restricting nonzero parameters to zero. Second, the nonzero parameters are as efficiently estimated as if the true subset of relevant variables is known. Thus, for the oracle properties to hold, selection consistency and asymptotic normality has to be satisfied as T goes to infinity.[14]

The asymptotic analysis follows the steps established in Song and Bickel (2011) and Lee and Liu (2012). Assume that the data Y_t are generated from an underlying model as in equation (2.1) where $U_t \sim \mathcal{N}(0, \Sigma)$ and that the PVAR model is stable. That is, all roots of $det(I_K - A_1 z - A_2 z^2 - ... - A_p z^p)$ are outside the unit circle.

Define the true parameter matrix as B^*. Assume that the covariance matrix is known. The inverse of the covariance matrix is denoted as Ω. If Ω is estimated consistently, it can be shown that the results derived in the following hold. The true coefficient in the k-th row and m-th column of B^* is defined as b^*_{km}. The vectorized true coefficient matrix is given by $b^* = vec(B^*)$. Let $J = \{(k, m) : b^*_{km} \neq 0\}$ denote the set of subscripts of nonzero parameters. The number of nonzero parameters, the cardinality of J, is given by $|J| = s$. The lasso estimator of b^*, as derived from the optimization problem in equation (2.3) under the $[1 \times NG^2p]$-vector of penalty parameters, λ, is denoted as $\hat{b}$. The b^*_J is the vector of true nonzero parameters with dimension $[s \times 1]$ and $\hat{b}_J$ is the estimator of b^*_J. Let $Z = I_K \otimes X'$, where X is the $[Kp \times T]$-matrix of right hand side lagged variables.

Define $a_T = \lambda_{km}$ for $k, m \in J$ and $c_T = \lambda_{km}$ for $k, m \notin J$. Assume that the lag length p can increase with growing T. Thus, λ_{km} is time dependent since it depends on p. The a_T is defined as the penalty term λ_{km} for a true nonzero parameter. The c_T gives the penalty term for true zero parameters. The specified penalty terms in *lassoPVAR* allow for different penalization of each variable. The introduction of time series and cross section penalty terms leads to stronger penalization of close to zero coefficients. Thus, the distinction of the penalty term in a_T and c_T is justifiable. Furthermore, the following assumptions are made:

(A1) $\Gamma := plim\ ZZ'/T$ exists and is nonsingular.

(A2) Nonzero parameters exist. The cardinality of J is nonzero, $|J| = s > 0$.

[13] For the definition of the oracle property, see, for example, Lee and Liu (2012) and Kock and Callot (2015).

[14] An increasing number of cross sections N increases the number of free parameters by adding equations and variables in each existing equation and not the number of time series observations.

(A3) Assume that $\sqrt{T}a_T \to 0$.

(A4) Assume that $\sqrt{T}c_T \to \infty$.

Thus, assumptions (A3) and (A4) require different rates of convergence properties for the penalty parameters associated with zero and truly nonzero coefficients.

Theorem 1. *Under the assumptions (A1) to (A4) the following results hold:*

(R1) ***Selection consistency****:* $plim \ \ \hat{b}_{km} = 0$ *if* $b^*_{km} = 0$.

(R2) ***Asymptotic normality****:* $\sqrt{T}(\hat{b}_J - b^*_J) \xrightarrow{d} \mathcal{N}(0, (\Omega \otimes \Gamma)^{-1}_J)$.

The proof of the theorem is provided in 2.D. The $(\Omega \otimes \Gamma)_J$ is the covariance matrix obtained by removing the row and column of $\Omega \otimes \Gamma$ corresponding to the elements $(k, m) \notin J$. Results (R1) and (R2) imply that if the penalty parameters satisfy the conditions given in (A3) and (A4), then *lassoPVAR* satisfies asymptotically the oracle properties. Theorem (R1) states the selection consistency. That is, for $T \to \infty$, a true zero parameter, b^*_{km} with $(k, m) \notin J$, is estimated consistently, meaning that, equaling zero. The second result, (R2), establishes the asymptotic normality for true nonzero parameters, b^*_{km} with $(k, m) \in J$.

2.3.5 Comparison to Other Estimation Procedures for PVAR Models

This section describes three further lasso specifications, the estimation of individual country VAR models, as well as the alternative existing estimation procedures for PVAR models in the literature, to which the performance of the lasso for PVAR models is compared. The alternative PVAR estimation methods are restricted least squares, the selection prior of Koop and Korobilis (2015b), and the cross-sectional shrinkage approach of Canova and Ciccarelli (2009).[15] As a general benchmark model, the PVAR model is estimated with ordinary least squared - this model is referred to as *OLS*. However, while it can serve as a benchmark for small models, *OLS* is unfeasible for larger models for which $T < Kp$.

Lasso with basic penalty. The first alternative lasso approach is a lasso with weighted sum of squared residuals as the loss function but without a penalty which explicitly captures panel properties. The time series penalty, α, is set to zero and the cross section penalty, c, equals one. Thus, the penalty parameter λ_{km} reduces to λ_k. In the following, this estimator will be referred to as *lassoVAR*.

Post lasso. Second, a post lasso is considered. The post lasso consists of two estimation steps, as explained by Belloni and Chernozhukov (2013). In the first step, a

[15]The two Bayesian approaches are only briefly described in this paper. See Koop and Korobilis (2015b), Canova and Ciccarelli (2004, 2009, 2013) for details.

lasso optimization problem is solved based on the proposed specification with weighted sum of squared residuals along with time series and cross section penalties. In the second step, the nonzero elements of the first step are re-estimated with OLS. Thus, the post lasso reduces the bias of the nonzero elements introduced via lasso shrinkage. This estimator is called *post lassoPVAR*.

Adaptive lasso. The third lasso alternative is the adaptive lasso, as proposed by Ren and Zhang (2010) for VAR models following the idea of Zou (2006). While the lasso shrinks all coefficients constantly depending on the penalty parameter, the adaptive lasso penalizes large nonzero coefficients less than very small coefficients. This is achieved by adaptive weights. Zou (2006) proposes weights, which are data-dependent, for the penalty parameter, $\hat{w}_{km} = \frac{1}{|b_{km}^{OLS}|^{\gamma}}$, where b_{km}^{OLS} is the OLS estimate and γ a constant. OLS estimates close to zero will increase the penalty parameter, leading to increased shrinkage, while large nonzero coefficients will decrease the penalty parameter. The adaptive lasso applied here, referred to as *adaptive lassoVAR*, uses the weighted sum of squared residuals and sets $\alpha = 0$ and $c = 1$. One issue of the adaptive lasso is the choice of the unbiased estimator for the weights. For very large models, OLS is not feasible if $T < Kp$. An alternative is to use ridge estimates or post lasso estimates as weights.[16]

The *lassoPVAR* allows for different penalty parameters for different coefficients. The specification of the time series and cross section penalties captures close to zero coefficients and penalizes those stronger. Consequently, *lassoPVAR* can be seen as an adaptive lasso.

Single-country VAR. This model assumes both a block-diagonal coefficient matrix and a block-diagonal covariance matrix. Hence, the model allows for no interdependencies across countries. Estimating the whole system is equal to an estimation of each single-country VAR model separately. The model is estimated with OLS. The estimator is called *single VAR*.

Restricted least squares. An estimation approach for PVAR models used in the literature, which is discussed here, is a restricted least squares estimation, called *restLS*. In this approach restrictions are set a priori. The here used restricted LS estimates a block-diagonal system ordering the variables in country blocks. Such a model assumes no dynamic interdependencies between countries. Setting the off-diagonal elements to zero reduces the number of free parameters. However, the assumption of no dynamic interdependencies between various economies is theoretically hard to justify. No restrictions are set on the covariance matrix.

[16]Compare with, for example, Kock and Callot (2015). However, using the post lasso will increase computation time while using ridge estimation requires further determination of hyperparameters.

Stochastic search specification selection. Another approach for estimating PVAR models is the Bayesian selection prior of Koop and Korobilis (2015b) called stochastic search specification selection (*SSSS*). The authors define weighted normal distributions as prior distributions that center around a restriction with a small or a large variance. Thus, the first part of the distribution shrinks the estimated parameter toward the restriction (small variance) while the second part allows for a more freely estimated parameter (large variance). Depending on a hyperparameter, which is set to be Bernoulli distributed, a parameter is drawn from the first or second part of the distribution. Koop and Korobilis (2015b) specify three different priors based on the possible restrictions: They search for no dynamic interdependencies, no static interdependencies and for homogeneity across coefficient matrices.

The prior centering around the no dynamic interdependency restriction is specified for an off-block-diagonal matrix of B of variables belonging to one country. The dynamic interdependency prior has the following form:

$$\begin{aligned} B_{ij} &\sim (1-\gamma_{ij}^{DI})\mathcal{N}(0,\tau_1^2 I) + \gamma_{ij}^{DI}\mathcal{N}(0,\tau_2^2 I) \\ \gamma_{ij}^{DI} &\sim Bernoulli(\pi^{DI}), \quad \forall j \neq i \end{aligned}$$

where B_{ij} is a off-block-diagonal matrix of B and $\tau_1^2 < \tau_2^2$. If $\gamma_{ij}^{DI} = 0$, B_{ij} is shrunk to zero, if $\gamma_{ij}^{DI} = 1$, B_{ij} is more freely estimated. Setting the prior on a block of variables of one country leads to a similar treatment of all variables of one country being either restricted (shrunk to zero) or not. The cross-sectional homogeneity prior is set on the diagonal coefficient matrices of the B matrix. The prior has the following form:

$$\begin{aligned} B_{ii} &\sim (1-\gamma_{ij}^{CSH})\mathcal{N}(B_{jj},\eta_1^2 I) + \gamma_{ij}^{CSH}\mathcal{N}(B_{jj},\eta_2^2 I) \\ \gamma_{ij}^{CSH} &\sim Bernoulli(\pi^{CSH}), \quad \forall j \neq i \end{aligned}$$

where B_{ii} and B_{jj} are block-diagonal matrices of B and $\eta_1^2 < \eta_2^2$. If $\gamma_{ij}^{CSH} = 0$, B_{ii} is shrunk to B_{jj}. Koop and Korobilis (2015b) specify a hierarchical normal mixture prior for the off-diagonal elements of the covariance matrix to build in no static interdependencies. Since no restrictions are set on the covariance matrix for the lasso solution and the forecast comparison is done on the reduced form, no restriction search for static interdependencies is done in the following exercises. The covariance is drawn from an inverse Wishart distribution. A Markov Chain Monte Carlo algorithm samples the estimated parameters as the posterior means.

Cross-sectional shrinkage approach. Another Bayesian estimation procedure for PVAR models is the cross-sectional shrinkage approach, *CC*, proposed by Canova and Ciccarelli (2004, 2009). Here, the parameters are factorized into common, country-specific, and variable-specific time-varying factors. Canova and Ciccarelli (2009) specify the model in a hierarchical structure:

$$\begin{aligned} b &= \Lambda F + e_t \\ Y_t &= Z_t \Lambda F + \epsilon_t \\ \epsilon_t &= U_t + Z_t e_t \\ e_t &\sim \mathcal{N}(0, \Sigma \otimes \sigma^2 I) \\ \epsilon_t &\sim \mathcal{N}(0, (I + \sigma^2 Z_t' Z_t)\Sigma) \end{aligned}$$

where Λ is a $[NG^2p \times f]$ matrix of loadings, F is an $[f \times 1]$ vector of factors, and $Z_t = I \otimes X_{t-1}$. Since the factors, F, are of a lower dimension than the vectorized B matrix, b, $f \ll NG^2p$ holds. The specified prior distributions for the covariance matrices are inverse Wishart and $b \sim \mathcal{N}(\Lambda F, \Sigma \otimes \sigma^2 I)$. The number of factors are N common factors for coefficients of each country, G common factors for coefficients of each variable, and one common factor for all coefficients.

An advantage of the approach is that it takes into account time variation. As one limitation, the cross-sectional shrinkage approach groups coefficients due to factorizing, however, it does not consider zero values in a specific way.[17] The procedure does not use possible sparsity for dimension reduction.

2.4 Simulation Studies

2.4.1 Simulation Set-Ups

The finite sample performance of the *lassoPAVR* is evaluated based on three Monte Carlo simulations. In the first simulation set-up data is generated from a stationary PVAR(1) model model that includes two countries and two variables per country. The number of time series observations is 100. The underlying PVAR model has the parameter matrix

$$A_1^{true} = \begin{bmatrix} 0.9 & 0.8 & 0 & 0 \\ 0 & 0.9 & 0 & 0 \\ 0.6 & 0.6 & 0.9 & 0 \\ 0.6 & 0.6 & 0.8 & 0.9 \end{bmatrix}$$

and normally distributed error terms, $U_t \sim \mathcal{N}(0, \Sigma^{true})$ with $\Sigma^{true} : \sigma_{ii} = 0.2$ and $\sigma_{ij} = 0.1$ for $i \neq j$. The PVAR model represents a scenario where the second country has no dynamic impact on variables of the first country while the first country's variables

[17]Korobilis (2016) elaborates further on this point.

affect the variables of country 2. This set-up could be a model including one big and one small economy, justifying the block of zeros in the upper part of the A_1-matrix. A second property of the model is that domestic variables have a greater impact than foreign variables have.

The number of parameters of this model is moderate. The coefficient matrix has 16 free coefficients out of which 10 are true nonzero coefficients. As a result, the methods aiming for dimension reduction, such as the lasso approaches and the two Bayesian procedures, are not able to provide substantial benefit by reducing the number of parameters to estimate. Rather, the simulation is conducted to analyze whether these methods perform comparable to standard OLS in terms of mean squared error and forecast accuracy.

In the second simulation data is generated from a stationary PVAR(4) with $U_t \sim \mathcal{N}(0, \Sigma^{true})$, $\Sigma^{true} : \sigma_{ii} = 0.2$, $\sigma_{ij} = 0.1$ for $i \neq j$, and $T = 100$. The model includes three countries and two variables per country. The set-up illustrates a larger and sparse model with parameter matrices

$$
\begin{aligned}
A_1^{true} &= \begin{bmatrix} 0.6 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0 & 0 \\ 0.4 & 0.4 & 0.6 & 0.5 & 0 & 0.4 \\ 0 & 0.4 & 0 & 0.6 & 0 & 0 \\ 0.4 & 0.4 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.4 & 0 & 0.6 \end{bmatrix}, \\
A_2^{true} &= \mathbf{0}, \quad A_3^{true} = \mathbf{0}, \\
A_4^{true} &= \begin{bmatrix} 0.35 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0.35 & 0 & 0 & 0 & 0 \\ 0.3 & 0.3 & 0.35 & 0.3 & 0 & 0.3 \\ 0 & 0.3 & 0 & 0.35 & 0 & 0 \\ 0.3 & 0.3 & 0.3 & 0 & 0.35 & 0 \\ 0 & 0.3 & 0 & 0.3 & 0 & 0.35 \end{bmatrix}.
\end{aligned}
$$

The model includes dynamic and static interdependencies as well as cross-sectional heterogeneities. It incorporates a time series pattern by lower coefficients for higher lags. Thus, the impact of a variable is smaller for lag four than for lag one. The second and third lag have no impact. This structure could be motivated by a model using quarterly data depicting seasonal patterns. In addition, foreign variables affect domestic variables less compared to the effect of domestic variables.

The second simulation provides a larger and sparser model than the model in simu-

Table 2.1: Summary of simulation set-ups

	Simulation (1)	Simulation (2)	Simulation (3)
lag length	1	4	4
number of countries	$N = 2$	$N = 3$	$N = 4$
number of variables	$G = 2$	$G = 2$	$G = 4$
number of nonzeros in B	10	34	432
number of elements in B	16	144	1024
fraction of nonzeros in B	62.50%	23.61%	42.19%

lation one. The coefficient matrices have 144 free parameters, out of which 34 are true nonzero coefficients, hence 23.61% of all coefficients of B are true nonzero coefficients. However, this model is still rather of medium size. The simulation enables analyzing whether efficiency gains compared to the benchmark OLS can already be found in medium sized models.

The DGP of the third simulation is based on a PVAR(4) with four countries and four variables per country. The U_t are normally distributed with $U_t \sim \mathcal{N}(0, \Sigma^{true})$, $\Sigma^{true} : \sigma_{ii} = 0.2$, $\sigma_{ij} = 0.1$ for $i \neq j$, and the length of the time series is $T = 100$. The coefficient matrices for lag $p = 1, ..., 4$ are lower triangular matrices where the diagonal elements are given by

$$(-0.8)^{(p-1)}0.8.$$

A column of the off-diagonal elements below the diagonal is given by

$$[0.5 - (p-1) \quad 0.5 - (p-1) \quad 0.5 - (p-1) \quad 0]'$$

repeated for each country. The coefficient matrices model that foreign lags are less important and that with increasing lag length the impact of the variables decreases. This large and sparse model allows for dynamic and static interdependencies as well as for heterogeneous coefficients across economies. In total, B has 1024 coefficients, of which 432 are nonzero coefficients, thus 42.19% are nonzero coefficients. The constant is set to zero in all three simulations without loss of generality since the data are standardized. Table 2.1 summarizes the simulation set-ups. The underlying models of simulation one and two are chosen to be all relatively small so that they allow the comparison to Bayesian PVAR model methods and least squares estimators. For simulation three some estimators are not feasible.

2.4.2 Performance Criteria

The performance of the lasso for PVAR models is evaluated along the following criteria.[18]

1. **Correct sparsity pattern**: The measure calculates how often the evaluated procedure takes the correct decision whether to include or exclude a variable. It measures how often are true relevant variables included and true irrelevant discarded averaged over all Monte Carlo replications.

2. **Fraction of relevant variables included**: It counts the number of true relevant variables included in the models relative to the number of all true nonzero coefficients averaged over all Monte Carlo replications.

3. **Number of variables included**: Reports the average number of variables included in the model. This measure evaluates the dimension reduction done by the estimator.

4. **MSE**: The mean squared error of the parameter estimates for one Monte Carlo replication is calculated as

$$MSE = \frac{1}{K^2 p} \sum_{k=1}^{K} \sum_{m=1}^{Kp} (\hat{b}_{km} - b_{km}^{true})^2$$

   where $\hat{b}_{km}$ is the estimate of the true parameter b_{km}^{true}. The MSEs are averaged over all Monte Carlo replications.

5. **MSFE**: The h-step ahead mean squared forecast error for one Monte Carlo replication is calculated as

$$MSFE = \frac{1}{T - h - T_2 - 1} \sum_{t=T_2}^{T-h_{max}} \left[\frac{1}{K} \sum_{j=1}^{K} (\hat{Y}_{j,t+h} - Y_{j,t+h})^2 \right]$$

   where $\hat{Y}_{j,t+h} = \hat{B}\hat{X}_{t+h_{max}-1}$ denotes the iteratively estimated h-step ahead forecast for t with $t = T_2, ..., T-1$ and $h = 1, ..., h_{max}$, $h_{max} = 12$. The MSFEs are averaged over t, over all variables and over all Monte Carlo replications.

Table 2.2 lists the estimators that are compared in the simulation studies. The OLS estimator serves as a benchmark. However, for larger models, where $T < Kp$, OLS is not feasible. The lag length of estimated PVAR models is set to the true lag length, which means one in the first simulation and four in the second and third simulations.

[18] Tibshirani (1996), Ren and Zhang (2010) or Kock and Callot (2015), for example, use similar criteria to assess the performance of the lasso.

Table 2.2: Overview of estimators

lassoPVAR	lasso for PVAR models with weighted sum of squared residuals, time series and cross section penalties, $\lambda_{km} = \lambda_k p^\alpha c$
lassoVAR	lasso for PVAR models with weighted sum of squared residuals, $\lambda_{km} = \lambda_k$, $\alpha = 0$ and $c = 1$
post lassoPVAR	post lasso for PVAR models: first step estimates lasso for PVAR models, with weighted sum of squared residuals, time series and cross section penalties, $\lambda_{km} = \lambda_k p^\alpha c$, second step re-estimates nonzero elements with OLS
adaptive lassoVAR	adaptive lasso for PVAR models with weighted sum of squared residuals, weights depend on OLS estimate, $\lambda_{km} = \lambda_k$, $\alpha = 0$ and $c = 1$
SSSS	selection prior of Koop and Korobilis (2015b)
CC	cross-sectional shrinkage approach of Canova and Ciccarelli (2009)
OLS	ordinary least squares estimation of PVAR model
restLS	restricted least squares estimation, block diagonal system on coefficient matrix, assumption of no dynamic interdependencies
single VAR	least squares estimation, block diagonal system for coefficient matrix and covariance, assumption of no dynamic and static interdependencies

2.4.3 Simulation Results

Table 2.3 and 2.4 contain the evaluation of the various estimation procedures along the five performance criteria for simulation one, marked as (1), simulation two, (2), and simulation three, (3). The first four columns present the results for the lasso techniques, the next two columns for the Bayesian methods, and the last three for the least squares estimators. The performance criteria are averages over 100 Monte Carlo replications.[19]

Overall, the simulation studies provide supporting evidence that the use of the lasso for PVAR models is beneficial in terms of lower mean squared errors and mean squared forecast errors relative to *OLS*. The forecast performance is additionally improved relative to the selection prior of Koop and Korobilis (2015b) and the factor approach of Canova and Ciccarelli (2009). Accounting for the panel characteristics in the penalty

[19]Further results for the simulations are given in 2.E.1.

Table 2.3: Performance evaluation of estimators

	lasso techniques				Bayesian methods		least squares		
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
Correct sparsity pattern in %									
(1)	55.13	54.69	55.13	53.06	-	-	37.50	37.50	37.50
(2)	40.83	54.54	40.83	51.43	-	-	34.72	34.72	76.39
(3)	47.91	51.96	47.91	51.52	-	-	39.06	39.06	57.81
Fraction of relevant variables included in %									
(1)	31.90	34.40	31.90	38.50	-	-	60.00	60.00	100.00
(2)	38.41	57.26	38.41	51.29	-	-	47.06	47.06	100.00
(3)	44.18	62.85	44.18	59.98	-	-	33.33	33.33	100.00
Number of variables included									
(1)	5.20	5.63	5.20	6.19	16	16	8	8	16
(2)	50.91	83.48	50.91	74.94	144	144	48	48	144
(3)	440.38	643.12	440.38	613.76	-	-	256	256	1024
Mean squared error relative to OLS									
(1)	0.9649	0.9654	0.9735	0.9578	0.9753	0.9631	0.9707	0.9700	-
(2)	0.5806	0.6426	0.6890	0.6335	0.8723	0.5755	0.5987	0.6232	-
(3)	0.2487	0.2819	-	0.2827	-	-	-	0.2394	-

Note: (1): Simulation 1, (2): Simulation 2, (3): Simulation 3. The correct sparsity pattern measures how often true relevant variables are included and irrelevant ones excluded. The fraction of relevant variables included counts the number of true relevant variables included in the models relative to the number of all true relevant variables. The number of variables included measures the dimension reduction. MSEs are relative to OLS. All measures are averaged over 100 Monte Carlo replications.

terms leads to better performance in terms of MSEs and MSFE relative to the *lassoVAR* which does not include time series or cross section properties in the penalty terms.

The *lassoPVAR* includes true relevant and discards irrelevant variables in 55.13% of all simulation draws of the first, in 40.83% of the second, and in 47.91% of the third simulation. The fraction of relevant variables included by *lassoPVAR* is 31.90%, simulation one, 38.41%, simulation two, and 44.18%, simulation three. The other lasso techniques reveal similar numbers while *restLS* and *single VAR* find the correct sparsity pattern in fewer cases but more often detect the fraction of relevant variables included. The number of detection of the correct sparsity pattern as well as the fraction of relevant variables included are low for all methods. The only exception, in some cases, is *OLS*. However, *OLS* does not reduce the dimension and, hence, is not feasible for larger systems.

The lasso techniques clearly reduce the dimension of the models. The *lassoPVAR* includes 32.50% of all variables in simulation one (number of variables included is on average 5.2), 35.35% in simulation two (50.91 variables included), and 43.01% in simulation three (440.38 variables included). That means that for (1) *lassoPVAR* includes fewer variables than the true number of nonzero coefficients, for (2) it picks too many variables, while for (3) it selects around the true number of nonzero coefficients. Hence, *lassoPVAR* performs best in the largest simulation with true nonzero coefficients around 40%. For model (2), which is the sparsest model, the performance of *lassoPVAR* is weaker. This might be due to the underlying model in simulation two, which sets the whole lags two and three to zero. This structure might be better captured by a model setting a whole block of coefficients to zero. These findings are also partly reflected in the numbers for the correct sparsity pattern and the fraction of relevant variables included.

The lower dimension reduction of *lassoVAR* compared to *lassoPVAR* might be driven by the specification of the penalty terms. The penalty terms of *lassoPVAR* introduce additional penalties on higher lags and foreign variables, which results in more variables excluded. *restLS* and *single VAR* reduce the number of variables by one-half in (1), one-third in (2), and one-fourth in (3). *SSSS* and *CC* are shrinkage approaches. Therefore, *SSSS* includes all variables. Since *CC* uses factors to reduce the number of parameters, the first three performance criteria are not applicable.

Compared to the benchmark *OLS*, all estimators reveal lower mean squared errors in all simulations. As expected, due to the moderate number of parameters in simulation one, the gain - measured in lower MSEs - of using lasso or the Bayesian methods is lower compared to the gain in the larger and sparser set-ups of simulations two and three. The MSEs, relative to OLS for simulation one, are in a range between 0.95 and 0.97 for all estimators. In simulation (2), *lassoPVAR* leads to a substantial reduction of 0.42 in the MSEs relative to OLS and performs second best compared to all other estimators. Only *CC* has a lower MSE at 0.5755. The *adaptive lassoVAR* and *post lassoPVAR* do not yield improvements compared to *lassoPVAR*. The fact that the second stage *OLS* estimation of *post lassoPVAR* relies on the possibly misspecified model of the first step of the lasso estimation could explain the performance of the *post lassoPVAR*. For simulation three some models are infeasible. The *lassoPVAR* clearly outperforms *OLS* with a MSE of 0.2487. Only *single VAR* has a slightly lower value, 0.2394. The weak performance of *OLS*, particularly in terms of MSE for the larger models, reflects the problem of overfitting.

The usage of the selection methods leads to a sizable reduction in mean squared forecast errors compared to *OLS* for all simulations, as shown in table 2.4. The presented one-step ahead, two-steps ahead, and six-steps ahead MSFEs are averaged over all t, all countries and variables and over the MC replications. The last three rows show the

Table 2.4: Mean squared forecast errors relative to OLS

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
MSFE for $h = 1$								
(1)	0.9541	0.9551	0.9606	**0.9511**	1.0170	0.9775	0.9531	0.9586
(2)	**0.7318**	0.7731	0.8183	0.7659	1.1948	0.7335	0.7390	0.7613
(3)	**0.1362**	0.1716	-	0.1724	-	-	-	0.1375
MSFE for $h = 2$								
(1)	0.9953	0.9953	0.9958	0.9953	1.0000	1.0000	**0.9948**	**0.9948**
(2)	**0.7707**	0.8170	0.8476	0.8117	1.2549	0.7754	0.7825	0.8056
(3)	**0.1468**	0.1874	-	0.1869	-	-	-	0.1477
MSFE for $h = 6$								
(1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2)	**0.9262**	0.9361	0.9410	0.9343	1.3866	0.9451	0.9316	0.9329
(3)	**0.0827**	0.0966	-	0.0967	-	-	-	0.0828
MSFE average over 12 horizons								
(1)	0.9957	0.9958	0.9964	**0.9955**	1.0014	0.9982	0.9956	0.9961
(2)	**0.9083**	0.9248	0.9351	0.9227	1.7136	0.9193	0.9136	0.9216
(3)	**0.0740**	0.0901	-	0.0899	-	-	-	0.0741

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true nonzero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true nonzero. (3): Simulation 3 - DGP of simulation 3 is generated from a four-country four-variable model with four lags, B has 1024 coefficients, 432 are true nonzero. MSFEs are relative to OLS and average over all t, all countries and variables and over 100 Monte Carlo replications.

MSFEs additionally averaged over 12 forecast horizons. Lowest MSFEs per row are marked in bold.

Even in the simulation with a small model, where dimension reduction is not required, MSFEs are lower for all estimators compared to *OLS*, except for *SSSS*, and are in a similar range compared among all estimators. For forecast horizon six, the estimators perform equally well. In the second simulation, the use of *lassoPVAR* improves the forecast accuracy for all horizons and produces the lowest MSFEs relative to all other methods. Hence, the results provide evidence that accounting for the inherent panel structure within the data by time series and cross section penalties pays off in terms of improved forecast accuracy. Averaged over 12 horizons, the MSFE is 0.9083, a gain of around 0.09 in forecast performance relative to *OLS*. The largest improvement is found for horizon one with a gain of around 0.27. The *lassoPVAR* also produces the lowest MSFEs in simulation three and substantially improves the forecast accuracy relative to *OLS*, with a MSFE averaged over 12 horizons of 0.0740.

For the covariance estimation of *lassoPVAR* the optimal selected ρ is equal to zero. Estimating the covariance with the two-step least squares procedure leads to similar performance results which can be found in 2.E.2.

2.5 Forecasting with Multi-Country Models

2.5.1 Forecasting Including a Global Dimension

This section assesses the forecasting performance of the PVAR model estimated with *lassoPVAR* for an empirical application. Of great interest for applied researches and policy makers are forecasts of macroeconomic variables. The forecasting exercise can shed light on whether forecasts of key macroeconomic variables of interlinked countries have to account for possible spillovers across countries. Since panel VAR models can exploit international interdependencies and commonalities, they are well suited as forecasting models including a global dimension.

Several studies stress the benefits of accounting for international dependences while forecasting national and international key macroeconomic variables. Ciccarelli and Mojon (2010) and Bjørnland et al. (2017) use a factor model for inflation and GDP forecasts. The authors report improved forecast performance when accounting for national and global factors. Koop and Korobilis (2015a) indicate that using a PVAR model, estimated by a factor approach, for forecasting key macroeconomic indicators of euro zone countries can lead to improvements in forecasts. Dees et al. (2007) forecast inflation of four euro area countries applying sectoral data. Their results provide evidence that forecasts with sectoral PVAR models outperform random walk or autoregressive models in the short run.[20]

2.5.2 Forecasting Applications

In this paper, forecast performance is evaluated for three different models, described in table 2.5. The benchmark model, model (1), includes monthly log changes in the harmonized index of consumer prices (CPI) and industrial production growth (IP) for five countries: Germany (DE), France (FR), Italy (IT), the United Kingdom (UK), and the United States (US). The second model extends the country set to ten countries by adding Denmark (DK), Greece (GR), Ireland (IE), Portugal (PT), and Spain (ES).

[20]Other papers use global VAR (GVAR) models to account for international linkages in forecasts. Pesaran et al. (2009), Greenwood-Nimmo et al. (2012), Dovern et al. (2016), Huber et al. (2016), and Garratt et al. (2016) provide evidence that GVAR models improve forecast performance relative to univariate benchmark models.

Table 2.5: Overview of empirical applications

	Model (1)	**Model (2)**	**Model (3)**
	$N = 5, G = 2, p = 6$	$N = 10, G = 2, p = 6$	$N = 6, G = 4, p = 6$
countries	DE, FR, IT, UK, US	DE, DK, ES, FR, GR, IE, IT, PT, UK, US	DE, ES, FR, IT, UK, US
variables	CPI, IP	CPI, IP	CPI, IP, REER, UN
T	131	131	131
NGp	60	120	144

Finally, the third model additionally uses the changes in unemployment rates (UN) and the real effective exchange rate (REER) for six countries: DE, ES, FR, IT, UK, and US.

The number of parameters per equation is larger than the number of observations for model (3) and very close to it for model (2). Hence, for these two models, *OLS* and estimators dependent on OLS, like *adaptive lasso*, *SSSS*, and *CC*, are not feasible. The data provided by the OECD cover the period from 2001:1 to 2016:6. All models include six lags.[21]

An out-of-sample forecast exercise is conducted. The forecasts are made for the period from 2011:7 to 2016:6. The up to twelve-horizons forecasts are iterated forecasts and are calculated by $\hat{Y}_{t+h} = \hat{B}\hat{X}_{t+h-1}$ for $h = 1, ..., 12$. The estimation is based on the data up to 2011:6 for the first forecasts and then rolling forward so that the same amount of time series observations is used for every forecast. That is, the estimated coefficient matrix, $\hat{B}$, used to calculate the forecast for 2011:7 is computed based on the various compared estimators using the observations from the start of the sample in 2001:1 to 2011:6. The next forecast, for 2011:8, is then based on data from 2001:2 to 2011:7. The choice of performing iterated rather than direct forecasts is motivated by the results of Marcellino et al. (2006), according to which iterated forecasts are preferred to direct ones despite theoretical findings demonstrating stronger robustness to model misspecification of the latter. The forecasts are evaluated by mean squared forecast errors. The forecasting performance of *lassoPVAR* is compared to the previously explained variants.

[21] The data are seasonally adjusted. Inflation is calculated as the log-differences of consumer price indices. UN is the difference of the unemployment rate from one period to the last period. The time series are stationary, de-meaned and standardized.

Table 2.6: Mean squared forecast error relative to OLS for model (1)

lasso techniques				Bayesian methods		least squares	
lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
MSFE for $h = 1$							
0.5708	0.5839	0.6166	**0.5825**	1.7722	0.6002	0.6114	0.6763
MSFE for $h = 2$							
0.5988	0.5978	0.6524	**0.5976**	1.7055	**0.5749**	0.6050	0.6453
MSFE for $h = 6$							
0.6985	0.7123	0.7372	0.7143	2.6418	**0.6624**	0.7153	0.7532
MSFE for $h = 12$							
0.7873	0.7951	0.8056	0.7953	4.5683	**0.7615**	**0.7775**	0.7790
MSFE average over 12 horizons							
0.6783	0.6869	0.7136	0.6870	2.8079	**0.6528**	0.6884	0.7155

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than 1 indicate better performance relative to OLS. *Average* are the MSFEs additionally averaged over all horizons.

2.5.3 Results of the Forecasting Exercises

Table 2.6 presents the averaged mean squared forecast errors relative to *OLS* for one-step, two-steps, six-steps, and twelve-steps ahead forecasts for model (1). Additionally, the last row indicates forecast performance averaged over twelve forecast horizons.[22]

First, the use of *lassoPVAR* improves forecast performance relative to *OLS*. The mean squared forecast error averaged over all countries, variables, t and horizons of *lassoPVAR* has the second lowest value with average MSFE of 0.6783. That means that on average using *lassoPVAR* for forecasting leads to a gain of 0.3217 in forecast accuracy compared to *OLS*. *lassoPVAR* produces stable forecasts over all twelve forecast horizons with MSFE relative to *OLS* in a range of 0.79 and 0.57. The benefit of using *lassoPVAR* relative to *OLS* is greatest for one-step ahead forecasts with a gain in forecast performance of 0.4292. None of the other estimators is statistically significantly better in terms of MSFEs than the *lassoPVAR*.[23]

Second, accounting for the time series and cross-sectional characteristics in the penalty terms leads to gains in the forecast accuracy. On average, *lassoPVAR* outperforms *lassoVAR* for all but one of the forecasts horizons. Third, the results provide evidence that the use of multi-country models compared to single-county models is

[22]Further results on country and variable basis are in 2.F.2.

[23]Results for the Diebold-Mariano Test assessing the statistical significance of the difference in MSFEs of the models are in 2.F.2.

Table 2.7: One-step ahead mean squared forecast error relative to OLS for model (1)

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.5755	0.5768	0.6122	0.5777	1.4021	**0.5612**	0.5943	0.6253
IP	**0.5661**	0.5910	0.6211	0.5872	2.1424	0.6392	0.6285	0.7273
Country specific mean squared forecast errors								
DE	0.5531	0.5846	0.6131	0.5908	1.0841	**0.5163**	0.5466	0.6234
FR	**0.6176**	0.6200	0.6809	0.6214	2.1145	0.6771	0.7055	0.7935
IT	**0.7265**	0.7563	0.7258	0.7485	2.3525	0.7998	0.8594	0.9906
UK	**0.5956**	0.6010	0.6119	0.5924	2.3022	0.6804	0.6064	0.6220
US	0.3613	0.3575	0.4515	0.3593	1.0080	**0.3273**	0.3389	0.3520
Mean squared forecast errors averaged over countries and variables								
Average	**0.5708**	0.5839	0.6166	0.5825	1.7722	0.6002	0.6114	0.6763

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than 1 indicate better performance relative to OLS. CPI denotes log differences of the consumer price index, while IP denotes the log differences of industrial production.

beneficial to improve forecast performance. MSFEs of *lassoPVAR* and *CC*, both models accounting for interdependencies across countries, are lower than for the *single VAR* model.

Table 2.7 presents disaggregated results providing variable and country specific one-step ahead mean squared forecast errors relative to *OLS*. The largest gain in forecast performance of the *lassoPVAR* is found for industrial production growth forecasts with a gain of 0.4339. *lassoPVAR* outperforms the other techniques for aggregated forecasts for FR, IT, and the UK. The mean squared forecast errors are particularly low for the US for selection methods compared to *OLS*. Variables of other countries have a low impact on US variables, thus, including these variables does not seem to improve the forecasts for the US.

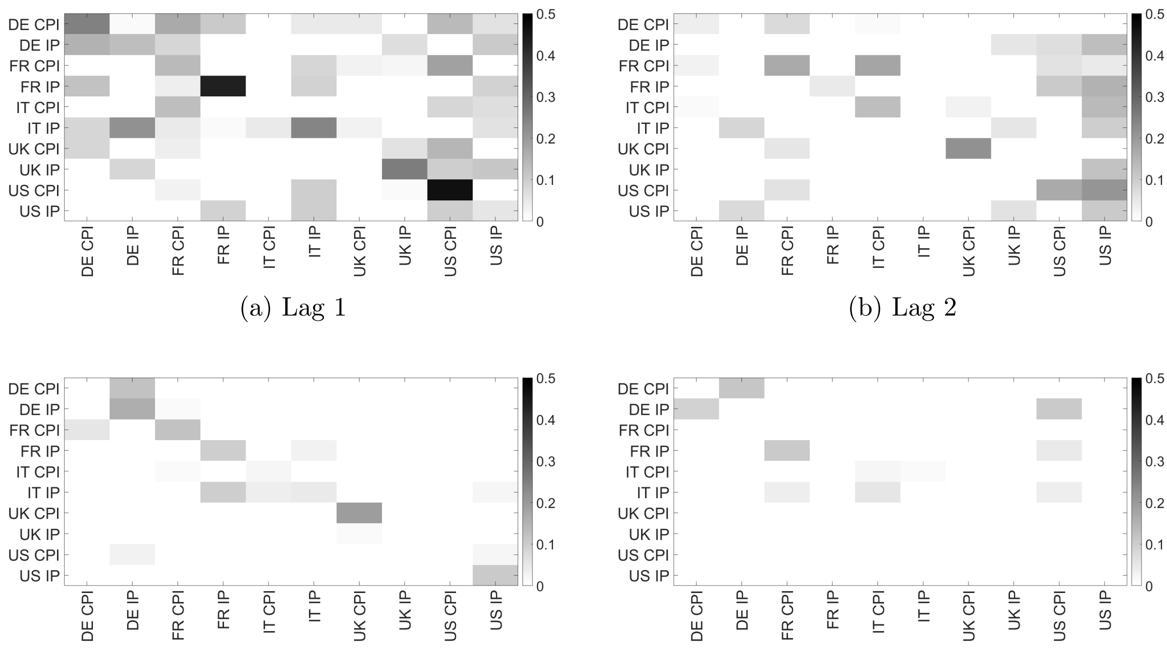
The results of the larger applications, model (2) and model (3), strengthen the findings. Since *OLS* is not feasible, table 2.8 compares MSFE of *lassoPVAR* and *single VAR* relative to the mean forecast. The mean forecast is calculated as the mean of the data used for the forecast. On average and for most of the horizons *lassoPVAR* outperforms the mean forecasts and the forecasts based on the single country model.

Table 2.8: Mean squared forecast error relative to mean forecast for model (2) and model (3)

$N = 10, G = 2, p = 6$		$N = 6, G = 4, p = 6$	
lassoPVAR	*single VAR*	*lassoPVAR*	*single VAR*
MSFE for $h = 1$			
0.9068	0.9807	**0.9948**	1.0402
MSFE for $h = 2$			
0.9540	1.0011	1.0476	1.0536
MSFE for $h = 6$			
0.9588	0.9764	**0.9333**	1.0104
MSFE for $h = 12$			
0.9519	**0.9253**	**0.9234**	0.9321
MSFE average over 12 horizons			
0.9526	0.9608	**0.9495**	0.9925

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to the mean forecast, MSFEs smaller than 1 indicate better performance relative to the mean forecast. *Average* are the MSFEs additionally averaged over all horizons.

Figure 2.1: Sparsity pattern of the coefficient matrix for model (1)



(a) Lag 1 (b) Lag 2

(c) Lag 3 (d) Lag 6

Note: Sparsity pattern of the coefficient matrix B. The graphs show the *lassoPVAR* estimates for lag 1, 2, 3 and 6 of all variables. Zero coefficients are colored in white, nonzero coefficients in gray shades depending on their values whereby negative values are multiplied by -1 and the darkest shade is given to the highest coefficient value.

CC shows good performance for small systems, but is infeasible for systems in which the number of parameters per equation exceeds the number of time series observations. That this is a relevant issue for applications, is shown in model (2) and (3) which are of reasonable or even still small size for models addressing potential macroeconomic questions in the context of international spillovers. A second issue with the factor approach is the difficulty mentioned for structural identification.

The sparsity pattern of the coefficient matrix for model (1) is given in figure 2.1.[24] The largest dynamic interdependencies across countries are found for the first and second lags. Own lags have the largest impact, as shown by the darker colors for diagonal elements. In addition, US variables affect variables of other countries, in particular for lags one and two. For benchmark model (1), 600 parameters of the coefficient matrix are estimated. The *lassoPVAR* reduces the dimension by setting 458 coefficients to zero. Thus, 23.67% of the estimated coefficients are nonzero elements.[25]

2.6 Conclusions

This paper develops a lasso technique for PVAR models, named *lassoPVAR*, and shows the asymptotic oracle properties of it. It specifies a penalized estimation problem using the weighted sum of squared residuals as the loss function and a penalty incorporating both time series and cross section properties. Thereby, it allows for an unrestricted covariance matrix, meaning that the estimation accounts for possible correlation between variables. The penalty term uses the inherent panel structure within the data. It specifies that more recent or domestic lags provide more information than more distant or foreign lags. As a result, a higher penalty is set for higher lags and foreign variables.

The main results of the paper are as follows. The *lassoPVAR* has the asymptotic oracle property meaning that selection consistency and asymptotic normality are established. Furthermore, the lasso for PVAR models achieves lower mean squared forecast errors, thus increasing forecasting performance compared to estimating the PVAR model with OLS. Compared to other Bayesian PVAR methods and single county models, the *lassoPVAR* improves forecasts, especially for larger models, while mean squared forecast errors are in a similar range for smaller models. These findings are supported by the simulation results and a forecasting exercise that includes up to ten advanced economies and up to four macroeconomic variables. Moreover, accounting for time series and cross section properties in the penalty term is beneficial for the forecast performance as *lassoPVAR* outperforms a lasso estimator without specific penalties. Additionally, the dimension reduction of the lasso techniques leads to reduced mean

[24]The sparsity pattern for lag 4 and 5 as well as for the covariance matrix are given in 2.F.2.

[25]The optimal penalty parameter ρ in the estimation of the covariance is selected to equal zero.

squared errors compared to OLS in the conducted simulations.

The method proposed in this paper may be of interest for applied researchers, since the lasso for PVAR models is able to deal with the curse of dimensionality problem in a multi-country model. *lassoPVAR* ensures the estimation feasibility by using the panel structure in the data and allows at the same time to include interdependencies and heterogeneities across countries in the model. The results presented show that the proposed lasso technique is a useful tool for estimating large PVAR models in practice.

However, the researcher must be aware that the performance of the lasso is sensitive to the suitability of the analyzed model for the penalized estimation technique. The lasso generally performs well in systems with a large number of parameters and existing sparsity. When few coefficients are large and the others close to zero, the lasso has usually low mean squared errors, while a good performance is not ensured for models deviating from these properties. This point is stressed by Hansen (2016) and is visible in the differences in results for the simulations with DGPs generated from a small and from a larger and sparse model. However, the benefit of the lasso for PVAR models is already visible through reduced mean squared errors and improved forecast accuracy in a simulation of moderate size with 165 parameters.

In future research, it may be interesting to further assess different specifications of the penalty term in the context of PVAR models. One possibility to capture the panel structure is the use of the group lasso, as proposed by Yuan and Lin (2006). The group lasso treats variables in groups, setting whole blocks to zero. This structure might be especially useful for analyses including smaller countries and globally more influential countries. Furthermore, variables in multi-country models might be highly correlated. This issue can be addressed with the elastic-net invented by Zou and Hastie (2005). This procedure is able to select groups of correlated variables while the lasso selects one variable out of a set of correlated variables.

Appendix

2.A The lasso Estimator

The optimization problem of the lasso for PVAR models minimizes over b_{km}. The b_{km} is the element of the B-matrix in the k-th row and m-th column. K is the number of countries times the number of variables, $K = NG$. The Y_j and X_m are of dimension $[1 \times T]$ for $j = 1, ..., K$ and $m = 1, ..., Kp$. The ω_{kj} denotes an element of the inverse of the covariance matrix, $\Sigma^{-1} = \Omega$. The λ_{km} is the penalty parameter and $|b_{km}|$ denotes the absolute value of b_{km}. The optimization problem is rewritten as

$$\begin{aligned}
\underset{b_{km}}{\text{argmin}} \quad \frac{1}{T} \Bigg[& \omega_{kk} \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right) \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right)' \\
& + \sum_{j \neq k}^{K} \omega_{kj} \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right) \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right)' \\
& + \sum_{j \neq k}^{K} \omega_{jk} \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right) \left(Y_k - b_{km} X_k - \sum_{i \neq m}^{Kp} b_{ki} X_i \right)' \\
& + \sum_{j \neq k}^{K} \sum_{l \neq k}^{K} \omega_{jl} \left(Y_j - b_{jm} X_k - \sum_{i \neq m}^{Kp} b_{ji} X_i \right) \left(Y_l - b_{lm} X_k - \sum_{i \neq m}^{Kp} b_{li} X_i \right)' \Bigg] \\
& + \lambda_{km} |b_{km}| + \sum_{j \neq k}^{K} \sum_{i \neq m}^{Kp} \lambda_{km} |b_{km}| .
\end{aligned}$$

This simplifies to

$$\begin{aligned}
\frac{1}{T} \Bigg[& \omega_{kk} \left(-2 b_{km} X_m Y_k' + b_{km} X_m X_m' b_{km} + 2 b_{km} X_m \sum_{i \neq m}^{Kp} X_i' b_{ki} + R_1 \right) \\
& + 2 \sum_{j \neq k}^{K} \omega_{jk} \left(-b_{km} X_m Y_j' + b_{km} X_m \sum_{i = m}^{Kp} X_i' b_{ji} + R_2 \right) + R_3 \Bigg] \\
& + \lambda_{km} |b_{km}| + \sum_{j \neq k}^{K} \sum_{i \neq m}^{Kp} \lambda_{km} |b_{km}| ,
\end{aligned}$$

where R_1, R_2 and R_3 collect the terms without b_{km}. Taking the derivative with respect to b_{km}:

$$\frac{1}{T}\left[\omega_{kk}\left(-2X_mY_k' + 2X_mX_m'b_{km} + 2X_m\sum_{i\neq m}^{Kp}X_i'b_{ki}\right)\right.$$
$$\left.+2\sum_{j\neq k}^{K}\omega_{jk}\left(-X_mY_j' + X_m\sum_{i=m}^{Kp}X_i'b_{ji}\right)\right] + sign(b_{km})\lambda_{km}$$
$$= 0.$$

Thus, b_{km}^{lasso} is equal to

$$b_{km}^{lasso} = sign\left(\frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j - \sum_{i=1}^{Kp}b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k - \sum_{i\neq m}^{Kp}b_{ki}X_i)X_m'}{X_mX_m'}\right)$$
$$\left(\left|\frac{\sum_{j\neq k}^{K}\omega_{jk}(Y_j - \sum_{i=1}^{Kp}b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'} + \frac{(Y_k - \sum_{i\neq m}^{Kp}b_{ki}X_i)X_m'}{X_mX_m'}\right|\right.$$
$$\left.-\frac{\lambda_{km}T}{2\omega_{kk}X_mX_m'}\right)$$

2.B Estimation of the Covariance Matrix

The covariance matrix is estimated using a graphical lasso (glasso) approach. Following Friedman et al. (2008) the subgradient of

$$log\ det(\Omega) - tr(S\Omega) - \rho\,||\Omega||$$

with respect to Ω is given by

$$W - S - \rho\Gamma = 0$$

with $W = \hat{\Sigma}$. The elements of Γ give the sign of each element of Ω by being either 1 or -1. For solving the glasso problem the partition

$$\begin{bmatrix} W_{11} & w_{12} \\ w_{12}' & w_{22} \end{bmatrix}\begin{bmatrix} \Omega_{11} & \omega_{12} \\ \omega_{12}' & \omega_{22} \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix}$$

is used. Here, W_{11} is the $(NG-1)\times(NG-1)$ block of W except the j^{th} row and column, w_{12} are the non-diagonal elements of the j^{th} column and row of W and w_{22} is the j^{th} diagonal element of W. The notation is the same for Ω. The partition of the matrix is done rotatively so that each j^{th} row and column is once ordered last. Now, to solve for w_{12} the subgradient is expressed as

$$w_{12} - s_{12} - \rho\gamma_{12} = 0$$
$$W_{11}z - s_{12} + \rho v = 0$$

where γ_{12} is the sign of ω_{12}, $z = -\frac{\omega_{11}}{\omega_{22}} = W_{11}^{-1}w_{12}$, $\gamma_{12} = sign(\omega_{12}) = sign(-\omega_{22}W_{11}^{-1}w_{12})$. Since $\omega_{22} > 0$, $sign(\omega_{12}) = -sign(z)$. The solution of the subgradient $\hat{z}$ gives than the value for w_{12} and $\omega_{12} = -\hat{z}\omega_{22}$. Since the diagonal elements of the covariance matrix are positive, $w_{ii} = s_{ii} + \rho \; \forall \; i$.

The glasso has the following three steps:

1. Set initial value $W = S + \rho I$. For diagonal elements $w_{ii} = s_{ii} + \rho \; \forall \; i$ do not update.

2. For each $j = 1, ..., NG$ update until convergence:
   a) Partition W and S.
   b) Solve $W_{11}z - s_{12} + \rho v = 0$.
   c) $w_{12} = W_{11}\hat{z}$.

3. Compute $\omega_{12} = -\hat{z}\omega_{22}$.

The optimal ρ is chosen over a grid of values by minimizing $BIC_\rho = log(\hat{\Sigma}_\rho) + \frac{log(T_1)}{T_1}df(\rho)$ as done similarly in Kock and Callot (2015). The degrees of freedom, $df(\rho)$, are the number of nonzero elements in $\hat{\Sigma}$. Since the penalty parameter ρ does not vary along the elements of the covariance matrix, the BIC criterion can be used which is faster than the cross-validation technique. The selection of the penalty parameter is done for the period up to T_1.

2.C Optimization Algorithm

The optimization problem is solved by a coordinate descent algorithm as proposed in Friedman et al. (2007) and Friedman et al. (2010). As a starting value B is set equal to a zero matrix. The covariance is estimated in the glasso step. The optimal penalty parameters are determined via a cross-validation technique minimizing MSFEs. The search of the optimal penalty parameters is done over a grid of values. The grid has a length of six for the simulations and twelve for the applications. This rather short length is due to the fact that using a finer grid increases computational time. The algorithm updates every element b_{km} for $k = 1, ..., K$ and $m = 1, ..., Kp$. The following steps are repeated until convergence is achieved. Update b_{km} as follows:

1. Calculate

$$\tilde{b}_{km} = \frac{(Y_k - \sum_{i\neq m}^{Kp} b_{ki}X_i)X_m'}{X_mX_m'} + \frac{\sum_{j\neq k}^{K} \omega_{jk}(Y_j - \sum_{i=1}^{Kp} b_{ji}X_i)X_m'}{\omega_{kk}X_mX_m'}$$

2. Set

$$\lambda_{km} = \begin{cases} \lambda_k p^{\alpha} c & \text{for foreign variables} \\ \lambda_k p^{\alpha} & \text{for domestic variables} \end{cases}$$

   where $\lambda_k > 0$, $\alpha > 0$, $c > 1$, and p is the lag length.

3. Calculate $\tilde{\lambda}_{km} = \frac{\lambda_{km}T}{2\omega_{kk}X_mX_m'}$

4. Calculate b_{km}^{lasso} as

$$b_{km}^{lasso} = \begin{cases} \tilde{b}_{km} - \tilde{\lambda}_{km} & \text{for } \tilde{b}_{km} > 0, \tilde{\lambda}_{km} < |\tilde{b}_{km}| \\ \tilde{b}_{km} + \tilde{\lambda}_{km} & \text{for } \tilde{b}_{km} < 0, \tilde{\lambda}_{km} < |\tilde{b}_{km}| \\ 0 & \text{for } \tilde{\lambda}_{km} \geq |\tilde{b}_{km}| \end{cases}.$$

5. Set the B-matrix of iteration $iter$ equal to values obtained in the last iteration, B_{iter-1}, that is $B_{iter} = B_{iter-1}$ for iteration $iter$.

Convergence is achieved when $max(|B_{iter} - B_{iter-1}|) < \epsilon$ where ϵ is a small number. The ϵ is chosen such that the $lassoPVAR$ converges to the OLS solution for a penalty parameter set to zero and weighted sum of squared residuals as the loss function. For the smaller simulation a conservative value of 0.0000001 is chosen, while for the large simulation (model 3) $\epsilon = 0.0001$.

2.D Proof of Selection Consistency and Asymptotic Normality

(R1) **Selection consistency**: $plim \ \hat{b}_{km} = 0$ if $b^*_{km} = 0$.

(R2) **Asymptotic normality**: $\sqrt{T}(\hat{b}_J - b^*_J) \xrightarrow{d} \mathcal{N}(0, D^{-1})$.

The vectorized true coefficient matrix is given by $b^* = vec(B^*)$. Let $J = \{(k,m) : b^*_{km} \neq 0\}$ denote the set of subscripts of nonzero parameters. The lasso estimator of b^* is denoted as $\hat{b}$. The b^*_J is the vector of true nonzero parameters with dimension $[s \times 1]$ and $\hat{b}_J$ is the estimators of b^*_J. Let $Z = I_K \otimes X'$ where X is the $[Kp \times T]$-matrix of right hand side lagged variables. The $y = vec(Y)$ and $u = vec(U)$ are a vector of dimension $[KT \times 1]$. The proof follows the line of arguments as in Song and Bickel (2011) and Lee and Liu (2012).

2.D.1 Proof of Asymptotic Normality

Let $\beta = \sqrt{T}(b - b^*)$. For the proof it is assumed that Ω is known. If $\hat{\Omega}$ is a consistent estimator of Ω, it can be easily shown that the same steps apply. The lasso optimization problem for the model $y = Zb + u$ is given by:

$$L(\beta) = \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right)' (\Omega \otimes I_T) \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right|$$

Using $\hat{\beta} = \underset{\beta}{\text{argmin}} L(\beta) = \underset{\beta}{\text{argmin}}(L(\beta) - L(0))$ it follows

$$\begin{aligned}
L(\beta) - L(0) &= \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right)' (\Omega \otimes I_T) \left(y - Z\left(b^* + \frac{\beta}{\sqrt{T}}\right)\right) \\
&\quad - (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right) \\
&= (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) - \left(Z \frac{\beta}{\sqrt{T}} \right)' (\Omega \otimes I_T)(y - Zb^*) \\
&\quad + \left(Z \frac{\beta}{\sqrt{T}} \right)' (\Omega \otimes I_T) \left(Z \frac{\beta}{\sqrt{T}} \right) - (y - Zb^*)'(\Omega \otimes I_T) \left(-Z \frac{\beta}{\sqrt{T}} \right) \\
&\quad - (y - Zb^*)'(\Omega \otimes I_T)(y - Zb^*) + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right) \\
&= \frac{1}{T} \beta' Z'(\Omega \otimes I_T) Z\beta - \frac{2}{\sqrt{T}} (y - Zb^*)' (\Omega \otimes I_T) Z\beta \\
&\quad + T \sum_{k=1}^{K} \sum_{m=1}^{Kp} \lambda_{km} \left(\left| b^*_{km} + \frac{\beta_{km}}{\sqrt{T}} \right| - |b^*_{km}| \right).
\end{aligned}$$

By assumption (A1) for $T \to \infty$

$$\begin{aligned}
\frac{1}{T} \beta' Z'(\Omega \otimes I_T) Z\beta &= \beta' \left(\Omega \otimes \frac{1}{T} Z'Z \right) \beta \\
&\to \beta' (\Omega \otimes \Gamma) \beta
\end{aligned}$$

and, since $u \sim \mathcal{N}(0, \Sigma \otimes I)$,

$$\begin{aligned}
\frac{1}{\sqrt{T}} (y - Zb^*)'(\Omega \otimes I_T) Z &= \frac{1}{\sqrt{T}} u'(\Omega \otimes I_T) Z \\
&\xrightarrow{\text{d}} \mathcal{N}(0, \Omega \otimes \Gamma).
\end{aligned}$$

Note that $\Omega = \Sigma^{-1}$ and

$$\begin{aligned} E\left(\frac{1}{\sqrt{T}}Z'(\Omega\otimes I_T)uu'(\Omega\otimes I_T)Z\frac{1}{\sqrt{T}}\right) &= \frac{1}{\sqrt{T}}Z'(\Omega\otimes I_T)E(uu')(\Omega\otimes I_T)Z\frac{1}{\sqrt{T}} \\ &= \frac{1}{T}Z'(\Omega\Sigma\otimes I_T)(\Omega\otimes I_T)Z \\ &= \frac{1}{T}(\Omega\otimes Z'Z)\rightarrow\Omega\otimes\Gamma. \end{aligned}$$

Under assumptions (A2) to (A4) the last term $T\sum_{k=1}^{K}\sum_{m=1}^{Kp}\lambda_{km}(|b^*_{km}+\frac{\beta_{km}}{\sqrt{T}}|-|b^*_{km}|)$ has the following asymptotic behavior for $T\rightarrow\infty$:

$$\begin{cases} \sqrt{T}\lambda_{km}(|b^*_{km}+\frac{\beta_{km}}{\sqrt{T}}|-|b^*_{km}|)\rightarrow 0 & \text{for } b^*_{km}\neq 0 \\ \sqrt{T}\lambda_{km}(|\beta_{km}|)\rightarrow\infty & \text{for } b^*_{km}=0 \end{cases}$$

since for $b^*_{km}=0$, it holds that $c_T\sqrt{T}\rightarrow\infty$. For $b^*_{km}\neq 0$, since $a_T\sqrt{T}\rightarrow 0$, it follows that $\sqrt{T}\lambda_{km}\rightarrow 0$ and $\sqrt{T}(|b^*_{km}+\frac{\beta_{km}}{\sqrt{T}}|-|b^*_{km}|)\rightarrow\beta_{km}sign(b^*_{km})$. As a result

$$L(\beta)-L(0)\xrightarrow{\text{d}}L(\beta)=\begin{cases} \beta'_J(\Omega\otimes\Gamma)_J\beta_J-2\beta_J D_J & \text{if } \beta_{km}=0\forall(k,m)\notin J \\ \infty & \text{if otherwise} \end{cases}$$

where β_J consists of $\beta_{km}\in J$ and $D_J\xrightarrow{\text{d}}\mathcal{N}(0,(\Omega\otimes\Gamma)_J$. The objective function $L(\beta)$ is minimized by

$$\hat{\beta}=\begin{cases} \hat{\beta}_J & =(\Omega\otimes\Gamma)_J^{-1}D_J \\ \hat{\beta}_{km} & =0\ \ \forall(k,m)\notin J \end{cases}$$

Thus, (R2) follows,

$$\hat{\beta}_J=\sqrt{T}(\hat{b}_J-b^*_J)\xrightarrow{\text{d}}\mathcal{N}(0,(\Omega\otimes\Gamma)_J)$$

2.D.2 Proof of Selection Consistency

For selection consistency to hold the probability that the coefficient estimate of a true zero parameter is different from zero converges to zero as T goes to infinity, $P(\hat{b}_{km}\neq 0)\rightarrow 0\ \ \forall(k,m)\notin J$. Suppose there is a $\hat{b}_{km}\neq 0$ for $(k,m)\notin J$. The Karush-Kuhn-Tucker conditions give the following:

$$0=\frac{\delta L(\hat{b})}{\hat{b}_{km}}+T\lambda_{km}sign(\hat{b}_{km}).$$

As shown by Song and Bickel (2011) for $T\rightarrow\infty$ the first term is dominated by the second. Since $c_T\sqrt{T}\rightarrow\infty$, the equation cannot equal zero. Thus, $P(\hat{b}_{km}\neq 0)\rightarrow 0$.

2.E Simulation

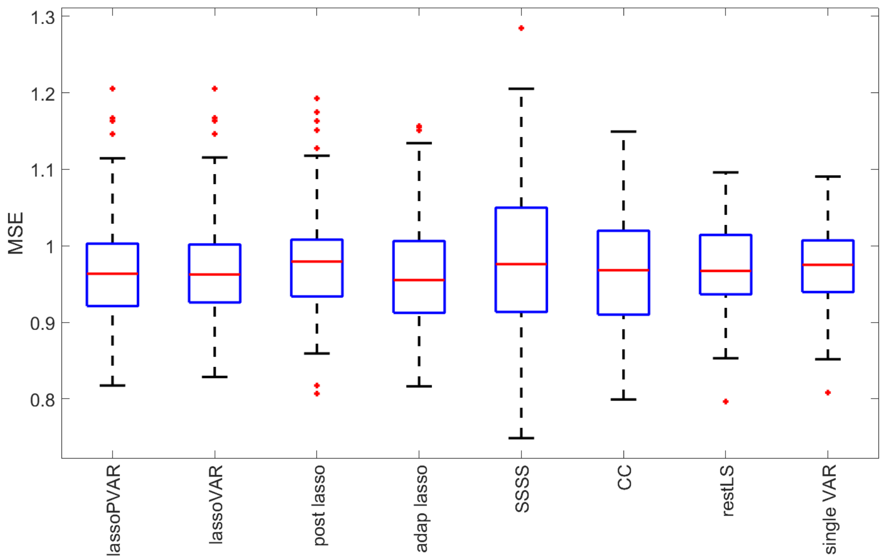
2.E.1 Additional Simulation Results

Table 2.9: Diebold-Mariano Test: test statistic and p-values

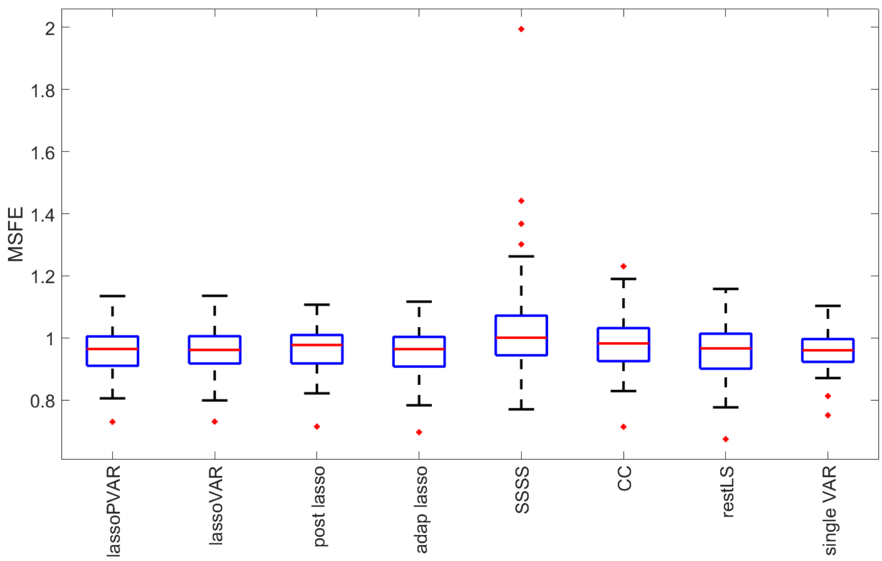
	lasso techniques			Bayesian methods		least squares		
	lasso VAR	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
horizon 1								
(1)	-0.37	-3.33	0.02	-6.53	-3.59	0.09	-0.74	-9.43
	0.36	*0.00*	*0.51*	*0.00*	*0.00*	*0.54*	*0.23*	*0.00*
(2)	-7.95	-9.81	-6.12	-17.90	-0.14	-0.86	-4.75	-15.56
	0.00	*0.00*	*0.00*	*0.00*	*0.45*	*0.20*	*0.00*	*0.00*
(3)	-16.86	-	-16.79	-	-	-	-0.70	-41.44
	0.00	-	*0.00*	-	-	-	*0.24*	*0.00*
horizon 2								
(1)	0.39	-2.69	0.33	-1.86	-1.58	0.78	0.97	-1.46
	0.65	*0.00*	*0.63*	*0.03*	*0.06*	*0.78*	*0.83*	*0.07*
(2)	-2.13	-2.19	-2.14	-2.12	-0.69	-1.41	-1.90	-2.13
	0.02	*0.01*	*0.02*	*0.02*	*0.25*	*0.08*	*0.03*	*0.02*
(3)	-2.11	-	-2.11	-	-	-	-0.29	-2.17
	0.02	-	*0.02*	-	-	-	*0.38*	*0.02*
horizon 6								
(1)	0.01	0.60	1.23	0.91	-0.52	1.99	1.14	1.03
	0.50	*0.73*	*0.89*	*0.82*	*0.30*	*0.98*	*0.87*	*0.85*
(2)	-1.07	-1.07	-1.10	-1.13	-1.11	-1.23	-1.13	-1.11
	0.14	*0.14*	*0.14*	*0.13*	*0.13*	*0.11*	*0.13*	*0.13*
(3)	-1.12	-	-1.13	-	-	-	-0.59	-1.12
	0.13	-	*0.13*	-	-	-	*0.28*	*0.13*

Note: (1): Simulation 1, (2): Simulation 2, (3): Simulation 3. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries and all MC draws.

Figure 2.2: Boxplots of MSEs and MSFEs relative to OLS for simulation 1



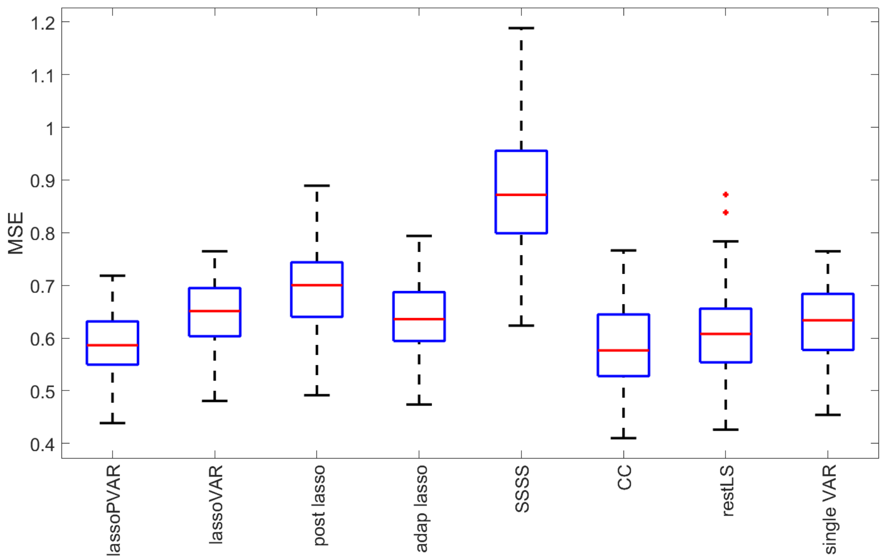
(a) Mean squared errors relative to OLS



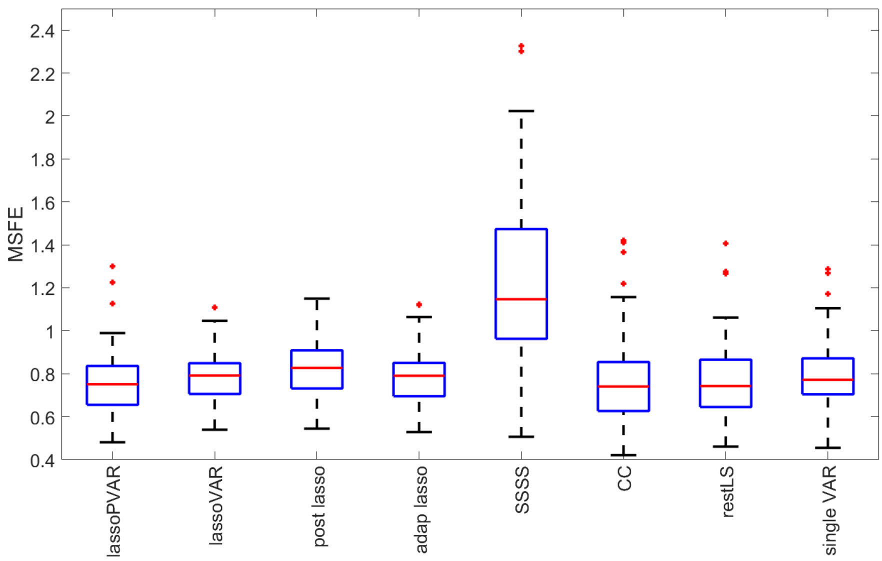
(b) One-step ahead mean squared forecast errors relative to OLS

Note: DGP of simulation 1 is generated from a sparse two-country two-variable model with one lag. (a) Boxplots show mean squared errors of estimates of B relative to OLS calculated as the average over the deviations of each $\hat{b}_{km}$ from the true value b_{km}^{true} for 100 replications. (b) Boxplots shows one-step ahead mean squared forecast error relative to OLS for 100 replications. MSFE is averaged over t and all variables.

Figure 2.3: Boxplots of MSEs and MSFEs relative to OLS for simulation 2



(a) Mean squared errors relative to OLS



(b) One-step ahead mean squared forecast errors relative to OLS

Note: DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags. (a) Boxplots show mean squared errors of estimates of B relative to OLS calculated as the average over the deviations of each $\hat{b}_{km}$ from the true value b_{km}^{true} for 100 replications. (b) Boxplots shows one-step ahead mean squared forecast error relative to OLS for 100 replications. MSFE is averaged over t and all variables.

2.E.2 Simulation Results for the Model with Covariance Estimated with OLS

Table 2.10: Performance evaluation of estimators: covariance estimated with OLS

	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*
Correct sparsity pattern in %				
(1)	55.31	54.94	55.31	53.37
(2)	37.05	52.57	37.05	48.76
Fraction of relevant variables included in %				
(1)	30.70	33.00	30.70	37.60
(2)	31.35	53.09	31.35	47.68
Number of variables included				
(1)	4.99	5.39	4.99	6.06
(2)	40.67	77.80	40.67	68.64
Mean squared error relative to OLS				
(1)	0.9632	0.9634	0.9692	0.9582
(2)	0.5711	0.6344	0.6512	0.6240

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true nonzero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true nonzero. The correct sparsity pattern measures how often true relevant variables are included and irrelevant excluded. The fraction of relevant variables included counts the number of true relevant variables included in the models relative to the number of all true relevant variables. The number of variables included measures the dimension reduction. All measures are averaged over 100 Monte Carlo replications.

Table 2.11: Mean squared forecast errors relative to OLS: covariance estimated with OLS

	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*
MSFE for $h = 1$				
(1)	**0.9550**	0.9555	0.9615	**0.9550**
(2)	**0.7299**	0.7733	0.7991	0.7627
MSFE for $h = 2$				
(1)	0.9953	0.9953	0.9953	0.9953
(2)	**0.7692**	0.8161	0.8276	0.8057
MSFE for $h = 6$				
(1)	1.0000	1.0000	1.0000	1.0000
(2)	**0.9232**	0.9317	0.9313	0.9300
MSFE average over 12 horizons				
(1)	0.9959	0.9959	0.9964	**0.9958**
(2)	**0.9065**	0.9232	0.9262	0.9198

Note: (1): Simulation 1 - DGP of simulation 1 is generated from a two-country two-variable model with one lag, B has 16 coefficients, 10 true nonzero. (2): Simulation 2 - DGP of simulation 2 is generated from a sparse three-country two-variable model with four lags, B has 144 coefficients, 34 are true nonzero. MSEs are relative to OLS. MSFEs are relative to OLS and average over all t, all countries and variables. All measures are averaged over 100 Monte Carlo replications.

2.F Forecasting Application

2.F.1 Penalty Parameters

Table 2.12: Grid values for penalty parameters - Application

	Model (1)	Model (2)	Model (3)
λ_k^1	0.376	0.528	0.6193
λ_k^2	0.3427	0.4809	0.5639
λ_k^3	0.3094	0.4338	0.5085
λ_k^4	0.2762	0.3867	0.4531
λ_k^5	0.2429	0.3396	0.3977
λ_k^6	0.2096	0.2925	0.3424
λ_k^7	0.1764	0.2454	0.287
λ_k^8	0.1431	0.1984	0.2316
λ_k^9	0.1098	0.1513	0.1762
λ_k^{10}	0.0765	0.1042	0.1208
λ_k^{11}	0.0433	0.0571	0.0654
λ_k^{12}	0.01	0.01	0.01

2.F.2 Additional Results of the Forecasting Exercises

The tables 2.13, 2.14 and 2.15 show the forecast evaluation split up into country and variable averages for two-steps ahead, six-steps ahead and twelve-steps ahead forecasts. Table 2.16 presents the average over all twelve forecast horizons. The differences in forecast performance along the two variables are exploited by averaging over all countries for each variable. The differences across countries are evaluated based on the MSFE averaged over the two variables.

lassoPVAR outperforms *OLS* for all variables for all horizons. The same holds for all countries. For one-step ahead forecasts *lassoPVAR* dominates the other estimators for IP and FR, IT, and the UK. For higher forecast horizons *CC* performs best in general. On average as for the six-steps ahead forecasts, *lassoPVAR* has the lowest MSFE for IP forecasts. For all horizons, forecast accuracy of the *lassoPVAR* is improved compared to *lassoVAR* for all countries and variables.

Table 2.13: Model (1): Two-steps ahead mean squared forecast error relative to OLS

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.6831	0.6936	0.7029	0.6924	1.7323	**0.6530**	0.6676	0.6987
IP	0.5145	0.5020	0.6019	0.5028	1.6788	**0.4968**	0.5423	0.5919
Country specific mean squared forecast errors								
DE	0.5570	0.5499	0.6342	0.5574	1.1186	**0.4938**	0.5380	0.6127
FR	0.6479	0.6352	0.7464	0.6391	1.9491	**0.6046**	0.6771	0.7256
IT	0.7256	0.7162	0.7846	0.7137	1.8606	**0.6898**	0.7822	0.8296
UK	0.6051	0.6276	0.6050	0.6186	2.5384	0.6755	**0.5821**	0.5822
US	0.4584	0.4603	0.4917	0.4592	1.0611	**0.4107**	0.4454	0.4764
Mean squared forecast errors averaged over countries and variables								
Average	0.5988	0.5978	0.6524	0.5976	1.7055	**0.5749**	0.6050	0.6453

Table 2.14: Model (1): Six-steps ahead mean squared forecast error relative to OLS

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7811	0.8054	0.7730	0.8070	2.9802	**0.6926**	0.7749	0.7961
IP	**0.6159**	0.6192	0.7015	0.6216	2.3033	0.6323	0.6556	0.7103
Country specific mean squared forecast errors								
DE	0.6315	0.6349	0.7076	0.6410	1.6743	**0.5625**	0.6466	0.7042
FR	0.7882	0.7950	0.8618	0.7986	3.8808	**0.7126**	0.8389	0.9034
IT	0.7810	0.7865	0.8369	0.7909	1.9969	**0.7648**	0.8561	0.9075
UK	0.7575	0.8015	**0.6934**	0.7964	3.6124	0.7688	0.7036	0.6908
US	0.5342	0.5435	0.5865	0.5444	2.0444	**0.5035**	0.5311	0.5602
Mean squared forecast errors averaged over countries and variables								
Average	0.6985	0.7123	0.7372	0.7143	2.6418	**0.6624**	0.7153	0.7532

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than 1 indicate better performance relative to OLS.

Table 2.15: Model (1): Twelve-steps ahead mean squared forecast error relative to OLS

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7792	0.7933	0.7891	0.7934	4.0400	**0.7063**	0.7686	0.7884
IP	0.7954	0.7970	0.8220	0.7972	5.0966	0.8166	0.7864	**0.7697**
Country specific mean squared forecast errors								
DE	0.8006	0.8077	0.8166	0.8084	4.2114	**0.7753**	0.7918	0.7904
FR	0.8124	0.8124	0.8628	0.8123	6.1899	**0.7802**	0.8138	0.8401
IT	0.7929	0.7994	0.8211	0.8001	5.0366	**0.7739**	0.8021	0.7880
UK	0.9751	0.9955	0.9204	0.9950	3.3784	0.9534	0.9274	**0.9201**
US	0.5556	0.5604	0.6068	0.5606	4.0252	**0.5245**	0.5524	0.5565
Mean squared forecast errors averaged over countries and variables								
Average	0.7873	0.7951	0.8056	0.7953	4.5683	**0.7615**	0.7775	0.7790

Table 2.16: Model (1): Average mean squared forecast error relative to OLS over all forecast horizons

	lasso techniques				Bayesian methods		least squares	
	lasso PVAR	*lasso VAR*	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*
Variable specific mean squared forecast errors								
CPI	0.7320	0.7478	0.7428	0.7478	2.6527	**0.6684**	0.7278	0.7517
IP	**0.6247**	0.6261	0.6845	0.6263	2.9632	0.6372	0.6490	0.6793
Country specific mean squared forecast errors								
DE	0.6290	0.6318	0.6828	0.6347	2.3950	**0.5851**	0.6298	0.6706
FR	0.7399	0.7413	0.7990	0.7417	3.7452	**0.7065**	0.7692	0.8065
IT	0.7339	0.7382	0.7742	0.7394	2.7701	**0.7059**	0.7935	0.8345
UK	0.7922	0.8214	0.7547	0.8167	3.1534	**0.7997**	0.7581	0.7544
US	0.4968	0.5020	0.5576	0.5026	1.9760	**0.4667**	0.4914	0.5116
Mean squared forecast errors averaged over countries and variables								
Average	0.6783	0.6869	0.7136	0.6870	2.8079	**0.6528**	0.6884	0.7155

Note: The forecast period is from 2011:7 to 2016:6. MSFEs are averaged over all t and are relative to OLS, MSFEs smaller than 1 indicate better performance relative to OLS.

Table 2.17: Diebold-Mariano test for model (1): Test statistic and p-values - relative to *lassoPVAR*

	lasso techniques			Bayesian methods		least squares		
	lasso VAR	*post lasso*	*adaptive lasso*	*SSSS*	*CC*	*restLS*	*single VAR*	*OLS*
$h = 1$	-0.68	-2.48	-0.82	-7.77	-0.12	-2.48	-0.72	-6.95
	0.25	*0.01*	*0.21*	*0.00*	*0.45*	*0.01*	*0.24*	*0.00*
$h = 2$	0.36	-2.31	0.27	-3.69	1.28	-1.36	-0.03	-3.82
	0.64	*0.01*	*0.61*	*0.00*	*0.90*	*0.09*	*0.49*	*0.00*
$h = 6$	-1.19	-1.75	-1.29	-2.10	1.53	-1.59	-0.70	-1.92
	0.12	*0.04*	*0.10*	*0.02*	*0.94*	*0.06*	*0.24*	*0.03*
$h = 12$	-1.17	-0.94	-1.23	-1.54	1.04	0.67	0.87	-1.61
	0.12	*0.17*	*0.11*	*0.06*	*0.85*	*0.75*	*0.81*	*0.05*

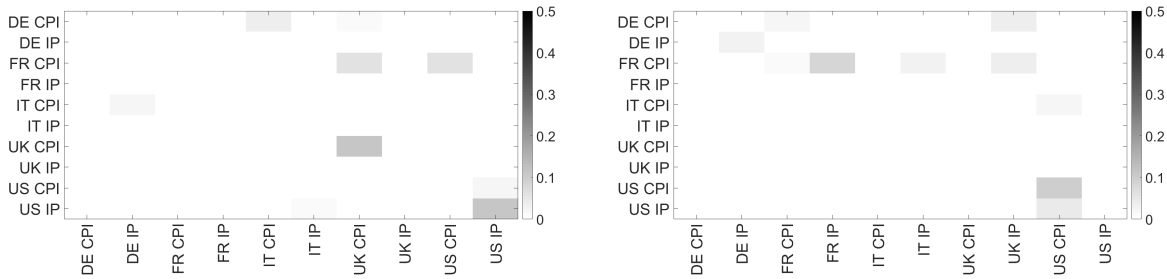
Note: The forecast period is from 2011:7 to 2016:6. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries.

Table 2.18: Diebold-Mariano test for model (2) and model (3): Test statistic and p-values - relative to *lassoPVAR*

	$N = 10, G = 2, p = 6$		$N = 6, G = 4, p = 6$	
	single VAR	*mean*	*single VAR*	*mean*
$h = 1$	-4.39	-1.03	-0.17	-0.12
	1.00	*0.00*	*0.43*	*0.45*
$h = 2$	-3.00	-0.18	0.62	0.65
	1.00	*0.09*	*0.73*	*0.74*
$h = 6$	-1.88	0.80	-1.90	-1.18
	0.99	*0.10*	*0.03*	*0.12*
$h = 12$	-1.36	1.44	-1.36	-0.55
	0.94	*0.16*	*0.09*	*0.29*

Note: The forecast period is from 2011:7 to 2016:6. Values of Diebold-Mariano test statistic and p-values which are presented in italic. MSFEs are compared to MSFEs of *lassoPVAR*. MSFEs are averaged over all variables and countries.

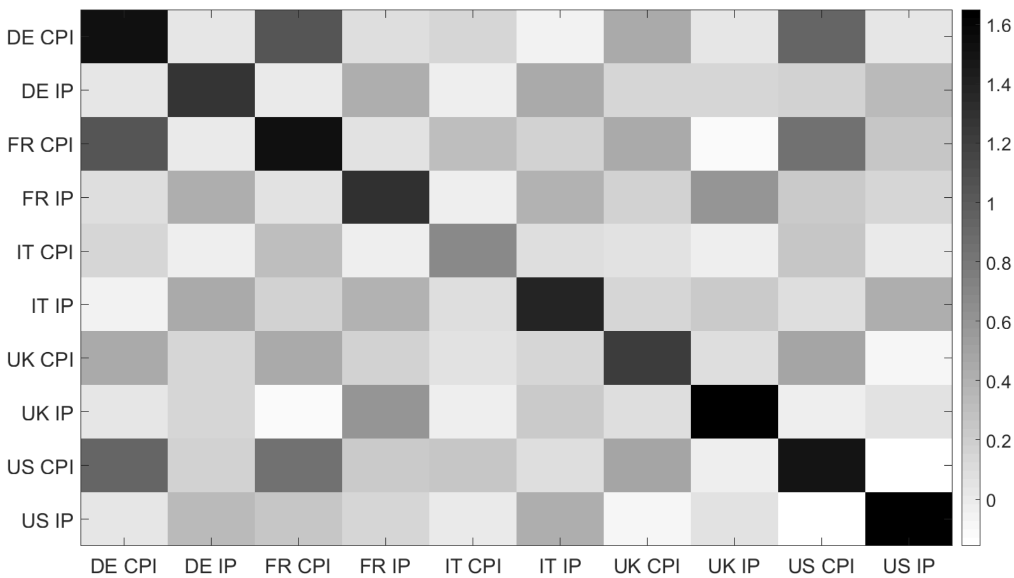
Figure 2.4: Sparsity pattern of the coefficient matrix for model (1): lag 4 and 5



(a) Lag 4 (b) Lag 5

Note: Sparsity pattern of the coefficient matrix B. Negative values are multiplied by -1.

Figure 2.5: Sparsity pattern of the covariance matrix for model (1)



CHAPTER 3

International Monetary Policy Transmission[1]

3.1 Introduction

The global financial crisis displayed the large risks of international contagion on financial markets. It showed how monetary policy can have international effects through exchange rate channels and foreign interest rates. In this light, it is key for policy makers and researcher alike to understand the cross-border effects of monetary policy. The interest in and relevance of this question is enhanced further in light of unconventional monetary policy measures, whose full effects are only slowly uncovered. In particular, an expansive monetary policy shock in one country may have cross-border effects due to exchange rate pressure and capital flows. Central banks may need to adjust their policy in reaction to foreign monetary policy action in order to reach their own inflation target. Thus, the effect of international monetary policy transmission on macroeconomic variables is of particular interest.

In this paper, we analyze empirically the degree of international transmission of monetary policy shocks of the United States, Great Britain, and the euro area (EA). The paper focuses on two main questions. First, are there macroeconomic effects of international monetary policy transmission and if so, what do they look like? Second, are there asymmetries across the US, UK, and EA regarding international monetary policy transmission? We use a Bayesian proxy three-country structural vector autoregressive model to trace the dynamic effects of conventional and unconventional monetary policy shocks on domestic and international macroeconomic variables. We estimate a three-country vector autoregressive (VAR) model including jointly variables for US, UK, and EA and use high-frequency data to identify separately one by one the monetary policy shocks of the US, UK, and EA. We successively augment the VAR, first, with a proxy

[1]This chapter is based on joint work with Gregor von Schweinitz. We gratefully acknowledge the financial support of the Leibniz Research Alliance "Crises in a Globalised World".

series constructed for the US to identify the US monetary policy shock, second, with a proxy for the UK for the identification of the UK monetary policy shock and, third, with a proxy for the EA to identify the EA monetary policy shock. In the following we use the term "country" also to describe the euro area.

The paper contributes to the literature, first, by analyzing not only cross-border effects of the US but also assessing effects of policy measures of the European Central Bank (ECB) and the Bank of England (BoE). Few papers analyze macroeconomic effects of international monetary policy transmission from central banks other than the Federal Reserve System (Fed). Kucharukov et al. (2016) and Bluwstein and Canova (2016) analyze spillovers from the ECB finding effects on exchange rates of European countries which are not part of the European Monetary Union and effects on output and inflation which vary across economies. Gerko and Rey (2017) find spillovers from the US to the UK but not vice versa. Rogers et al. (2014, 2017) provide evidence for financial spillovers for monetary policy shocks of the UK, Japan, and the EA.

Second, the paper complements the extensive literature on financial spillovers with a focus on macroeconomic effects using a three-country model accounting for linkages across the three economies. Recent studies such as Hausman and Wongswan (2011), Glick and Leduc (2012), Bauer and Neely (2014), Neely (2015), and Fratzscher et al. (2018) provide evidence that conventional and unconventional monetary policy of the Fed has cross-border financial spillover effects.

Third, we use external instruments to identify one monetary policy shock for each country in a multi-country VAR model. Structural identification in multi-country models is a challenge and usually solved with recursive identification where the order of variables matters. Papers using two-country VAR models for studying monetary policy transmission mostly base their identification on block-exogeneity. These assumptions are hard to justify when focusing on countries other than the US or when adding more countries. Chen et al. (2016), Kucharukov et al. (2016), and Bluwstein and Canova (2016) are recent examples for studies analyzing international monetary policy spillovers based on recursive identification schemes. External instrument identification allows us to identify similar shocks for different countries. We use changes of forward rates around central bank announcements to identify monetary policy shocks in VAR models (see e.g., Faust et al., 2003, 2004; Gertler and Karadi, 2015; Cesa-Bianchi et al., 2016; Miranda-Agrippino, 2016; Caldara and Herbst, 2016; Hachula et al., 2016; Gerko and Rey, 2017; Rogers et al., 2017). At the core of this identification lies the assumption that forward rates embody all expectations about a change of central bank policy until maturity. If rates change in a short window around central bank announcements, this change has to come from a change in expectations, i.e., an unexpected monetary policy shock.

The model with the identified US monetary policy shock uses monthly data from January 1973 to June 2017, the VAR model with the identified UK monetary policy uses data from January 1982 to March 2017, and for the VAR model with the identified EA monetary policy shock data range from January 2000 to October 2017. To analyze monetary policy shock transmission the VAR models include variables such as a monetary policy indicator, consumer price index, industrial production, unemployment rate, credit spread, stock price index, long term government rate, lending rate, and exchange rates. The external instruments used for identification of the monetary policy shocks measure conventional and unconventional monetary policy. We take the proxy of Gertler and Karadi (2015) for the US, the series of Gerko and Rey (2017) for the UK, and of Rogers et al. (2017) for the EA.

We estimate a Bayesian proxy three-country structural VAR model following the procedure of Caldara and Herbst (2016) but expanding it to the three-country dimension. A challenge is to ensure the estimation feasibility of the rather large VAR model. We use that the model includes three countries and adjust a Bayesian selection prior of Korobilis (2013) such that it utilizes the panel structure similarly as proposed in chapter 1. The prior searches for zero restrictions on lagged foreign variables and thereby reduces the model dimension. Our prior works under the hypothesis that domestic lags are more important than foreign lags and not all foreign lags have an impact on domestic variables. The Bayesian selection prior with a panel structure determines in a data-driven way which foreign lags can be set to zero. The posterior distribution sets zero coefficients on some foreign lags and is thus capable to estimate a rather large three-country VAR model. In short, we follow approaches that are also employed to estimate panel vector autoregressive (PVAR) models (see chapter 1).

Our results show asymmetries regarding the importance of foreign monetary policy for domestic variables. A US monetary policy shock spills-over to the UK and EA causing an appreciation of the exchange rates, while a UK monetary policy shock affects the British-Pound-Euro exchange rate but does not impact the exchange rate to the US dollar. Exchange rates do not respond to a monetary policy shock of the EA. We find no evidence for substantial macroeconomic effects. Our findings do not suggest the need for policy actions of central banks to prevent harmful macroeconomic spillover effects caused by foreign monetary policy shocks. However, the insignificance of our findings in the multi-country models might be explained by the shock identification and the high estimation uncertainty.

In the following, section 3.2 relates the paper to the previous literature. Section 3.3 briefly describes the monetary policy frameworks of the Fed, the Bank of England

and the European Central Bank and the transmission channels of monetary policy spillovers. In section 3.4 the Bayesian proxy three-country VAR model as well as the identification strategy are introduced. Section 3.5 provides an overview of the data and in section 3.6 the results are presented. Finally, section 3.7 concludes.

3.2 Literature

The paper contributes to the literature, first, by focusing not only on the effects of US monetary policy transmission but also on the UK and EA. Second, we analyze international macroeconomic effects of conventional and unconventional monetary policy shocks using a three-country model that allows to capture interlinkages and heterogeneities across the US, UK, and EA. Third, we do not rely on any block-exogeneity assumption or recursive identification but use external instruments to identify the monetary policy shocks of the three central banks.

Our first contribution expands the literature on spillovers from the UK and EA. Some recent studies focus on international macroeconomic effects of monetary policy shocks other than the US. Kucharukov et al. (2016) and Bluwstein and Canova (2016) analyze spillovers of the ECB's unconventional monetary policy to European Union countries which are no members in the European Monetary Union. Both studies base their identification of the monetary policy shock on a recursive structure or block-exogeneity assumptions. Kucharukov et al. (2016) focus on the transmission of conventional and unconventional monetary policy of the ECB using two-country VAR models. They construct a synthetic indicator of EA monetary conditions to measure the monetary policy of the ECB. The authors find effects on inflation and output for conventional monetary policy transmitted by exchange rates and interest rates while unconventional monetary policy mainly affects exchange rates and only causes output responses for some countries. Using two-country Bayesian mixed frequency VAR models Bluwstein and Canova (2016) study spillovers of EA unconventional monetary policy measured based on balance sheet variables. EA unconventional monetary policy shocks cause positive or insignificant output reactions while inflation responds for some countries negatively and for some positively. Spillovers occur via an exchange rate channel and financial channels.

Furthermore, Gambacorta et al. (2014) use a panel VAR model to study cross-border effects of conventional monetary policy of eight advanced economies identified with sign and zero restrictions as an exogenous innovation to the central bank balance sheet. They focus on domestic effects and only allow for correlations among residuals of the same endogenous variable across economics. Output responses are similar and positive across countries while prices respond differently. Gerko and Rey (2017) use external instruments to identify US and UK monetary policy shocks in a VAR model. They differentiate between effects of movements in the policy instruments and forward guidance. The authors provide evidence for spillovers from the US monetary policy to the UK while they find no evidence for spillovers from the UK to the US. A tightening in the US monetary policy causes an increase in the UK mortgage spread and leads to an appreciation of the British pound.

Few studies focus on financial spillover effects of monetary policy from different central banks than the Fed. Rogers et al. (2014) use intraday changes in government bond yields around policy announcements for the US, UK, EA, and Japan. Their regression results show that foreign bond yields are affected. The effects are strongest for the US monetary policy to non-US yields. Fratzscher et al. (2016) focus on unconventional monetary policy of the ECB and find that financial markets in the euro area are mainly affected. In their event study they show evidence for increased equity prices, lower credit risk and negligible effects on international yields. Rogers et al. (2017) use external instruments in two-country VAR models to identify unconventional monetary policy shocks for the US, UK, EA, and Japan and to assess effects on domestic and foreign interest rates and exchange rates. The study does not analyze effects on foreign real variables and of non-US shocks on non-US variables. The dollar depreciates in response to an expansionary US monetary policy shock disentangled into a target surprise, forward guidance surprise and asset purchase surprise. Foreign short term and long term interest rates decrease. US interest rates react to UK and Japanese monetary easing positively while the ECB easing shock causes a raise in US interest rates.

With our second contribution, focusing on international macroeconomic effects of monetary policy shocks in a three-country model, we relate to the literature on international monetary policy transmission. International monetary policy spillovers are to a great extent analyzed in two-country VAR models focusing on the effect of conventional monetary policy of the US. These models do not account for global interlinkages besides bilateral linkages. Furthermore, the structural identification of monetary policy is mainly based on block-exogeneity assumptions. Thus, these models are to a great extent used to analysze the effect of US monetary policy on a small country in order to justify the identifying assumptions. Two-country VAR models carry the risk that identified shocks mix a national and an international component, where the latter is not explicitly considered as international dependencies are not part of the model. Georgiadis (2017) points out this issue for identification which is not based on additional exogenous series such as narrative approaches or high-frequency identification.

Several studies such as Kim and Roubini (2000), Kim (2001), Canova (2005), and Maćkowiak (2007) provide evidence that conventional US monetary policy shocks are transmitted globally affecting advanced and emerging markets mainly through interest rate spillovers. Chen et al. (2016) study macroeconomic effects of quantitative easing of the Fed in a global vector error correction model for seventeen advanced economies and emerging market economies (EME). The authors show that cross-border spillovers vary across economies and are larger for emerging markets. They find positive effects on output growth and inflation and an appreciation pressure for EME. Advanced economies respond to a negative monetary policy shock of the US by loosening their monetary policy while EME respond in various ways. Their identification is based on a recursive scheme.

Several recent studies focus on international financial spillovers of monetary policy shocks of the US. Our paper extends this literature by also assessing real effects. The papers on international financial spillovers rely on an alternative identification strategy and use high frequency data for the identification of the monetary policy shocks in event study approaches. They provide evidence for large international financial spillovers from conventional and unconventional US monetary policy. Hausman and Wongswan (2011) focus on the effect of conventional US monetary policy on foreign equities, interest rates, and exchange rates. They find that for the 49 analyzed countries equity indexes mainly react to target surprises while exchange rates and long term interest rates respond to a path surprise capturing news about the revision in the future policy path. Using an event study approach Glick and Leduc (2012) analyze the impact of asset purchase programs of the US and the UK on financial asset prices of advanced economies. While the US shock causes long-term interest rates to decrease internationally, the UK shock has negligible effects in magnitudes. Both monetary policy shocks lead to a fall in the respective currency. Bauer and Neely (2014) find evidence that the asset purchases of the Fed reduce Canadian long-term yields, while for Australia, Germany and Japan portfolio balance effects are visible. Neely (2015) shows that unconventional monetary policy of the Fed lowers long-term yields internationally and leads to a depreciation of the US dollar. Fratzscher et al. (2018) analyze the effects of US quantitative easing on portfolio flows providing evidence for strong impacts on inflows to emerging market economies.

With our third contribution, using external instrument identification in a multi-country model, we relate to the literature on using external instrument identification with high-frequency data for monetary policy shocks in VAR models. Gertler and Karadi (2015) use surprises in Fed funds futures to identify US monetary policy shocks in a VAR model. Cesa-Bianchi et al. (2016) provide an instrument for UK monetary policy shocks using changes in 3-month sterling futures contracts. Miranda-Agrippino (2016) proposes to project raw monetary policy surprise series onto a set of central banks' forecasts in order to assure the exogeneity of the monetary surprises. She provides series for the US and UK. Caldara and Herbst (2016) use a Bayesian Proxy VAR model to identify US monetary policy shocks with high frequency data. Hachula et al. (2016) identify unconventional monetary policy surprises of the ECB with the high frequency identification in a VAR model.

3.3 Cross-Border Transmission of Monetary Policy

3.3.1 Monetary Policy in the United States, United Kingdom, and the Euro Area

The central banks of the US, UK, and EA all follow slightly different monetary frameworks. In particular, meeting schedules and the information flow regarding monetary policy decisions and the underlying now- and forecasts of central banks differ. As these differences are important for our identification strategy of monetary policy shocks, we briefly describe the monetary framework of the three central banks.

The monetary policy of the Fed is set by the Federal Open Market Committee (FOMC). The FOMC holds eight regularly scheduled meetings per year, plus additional meetings when needed.[2] After every meeting, there is a press release, which contains (a) the decision on the monetary policy stance (i.e., an interest rate decision), and (b) the macroeconomic forecasts on which the FOMC decision is based. That is, the press release provides markets with news on target interest rates. In particular, an increase of federal fund futures from shortly before the meeting to shortly after the press release can be seen as a tightening monetary policy shock.[3] In addition to a monetary policy shock, the press release conveys additional information on expected future macroeconomic developments. Jarociński and Karadi (2017) use this differentiation to solve the puzzle that stock markets do not necessarily react to interest rate surprises. Three weeks after every FOMC meeting, the Fed releases minutes. However, there is wide agreement in the literature that these minutes do not contain significant additional information for markets that can be employed to identify monetary policy shocks.

The federal funds rate reached its effective lower bound by November 2008. At the same time, the Fed set up a first large-scale asset purchase program. In this period

[2] https://www.federalreserve.gov/monetarypolicy/fomccalendars.html

[3] This interpretation entails that the FOMC decided on higher target interest rates than previously expected.

of unconventional monetary policy, the FOMC also put a stronger focus on forward guidance (i.e., providing the public with additional information on expectations).

The Monetary Policy Committee of the Bank of England meets monthly to set target interest rates and the monetary policy stance. After the meeting (on either the first or second Thursday of the month), there is a press release which only communicates the new target interest rate. That is, contrary to the Fed, press releases after meetings only can trigger a monetary policy shock. Meeting minutes are reported on the Wednesday two weeks after every meeting. These communications in turn provide information on central bank expectations. In addition, the BoE releases an inflation report commenting on the current and prospective economic developments. This report is published four times a year on the second Wednesday of February, May, August and November. Gerko and Rey (2017) argue that the clear separation of monetary policy shocks from central bank information across time implies that shock identification is cleaner in the case of the UK.

The BoE needed to conduct unconventional monetary policy from March 2009 onwards, when its policy rate was reduced to 0.5%. At that time, it also set up an asset purchase program. Forward guidance was, however, only introduced in August 2013.

The Governing Council of the European Central Bank meets every two weeks and decides on its monetary policy every six weeks.[4] Its decision is directly communicated in a press release and conference. An account of the Governing Councils's meetings is published four weeks after each monetary policy meeting. The information content of the different parts of communication of the ECB are comparable to those of the Fed. That is, the majority of new and relevant information is already released in the press release and during the press conference accompanying the monetary policy decision.

The ECB reduced the policy rate by 50 basis points on 8 October 2008. It subsequently started with the introduction of unconventional monetary policy actions by introducing the Longer-Term Refinancing Operations, the Securities Market Programme, and the Outright Monetary Transaction.

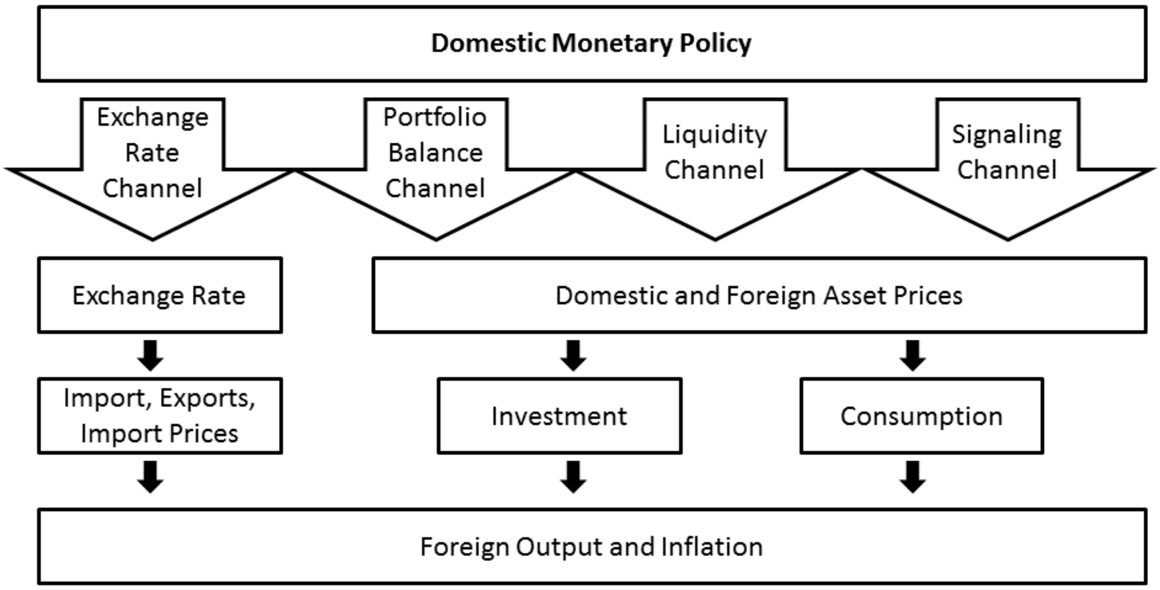
3.3.2 Transmission Channels

Monetary policy does not only have domestic effects, but can also affect foreign economies through a multitude of channels. The most important of these are the *exchange rate channel* and different *financial channels*, which are summarized in Figure 3.1.

The exchange rate channel can affect trade and import prices. Thus, it can cause macroeconomic spillovers across borders by stimulating a reaction of foreign output and

[4] https://www.ecb.europa.eu/press/govcdec/html/index.en.html and https://www.ecb.europa.eu/ecb/orga/decisions/govc/html/index.en.html

Figure 3.1: International monetary policy transmission



inflation. A monetary policy easing in the domestic country boosts domestic output and consumption. Thus, it also increases imports and thereby affects foreign output positively. However, this demand effect is counteracted by a substition effect. As domestic interest rates fall and expected inflation increases, the exchange rate depreciates. This implies an expenditure-switching effect since domestic goods become cheaper in comparison to foreign goods. The substitution from imports to domestic output dampens foreign economic development. The aggregate transmission of domestic monetary policy via the exchange rate channel depends on the relative strength of the demand and the substitution effect.

The exchange rate channel can be affected by potential policy coordinations across economies. Obstfeld and Rogoff (2002) argue that international spillovers of monetary policy are unintended, which implies that gains from policy coordinations may in practice only be small. Taylor (2013), however, argues that foreign monetary policy actions can trigger a pressure to deviate from planned domestic monetary policy in order to prevent exchange rate movements.

While for conventional monetary policy the exchange rate channel is the main transmission mechanism, the period of zero interest rates and unconventional monetary policy triggered a vast literature that addresses the importance of various financial spillover channels.[5] Unconventional monetary policy can have international effects via the *portfolio balance channel*, the *liquidity channel*, and the *signaling channel*. Large

[5]See for example Glick and Leduc (2012), Bauer and Neely (2014), Neely (2015), Fratzscher et al. (2016), and Fratzscher et al. (2018) for a detailed description of the various financial channels.

asset purchases programs were targeted to decrease long-term yields, improve financial conditions and thus stimulate economic growth. The portfolio balance channel passes on domestic monetary policy by increased asset purchases which decrease long-term yields as the supply of long-term securities is reduced. Investors in turn purchase both alternative domestic and foreign assets as substitutes. This leads to an increase in foreign asset prices boosting foreign consumption, investment and output.

The liquidity channel works through improved financial conditions directly. An asset purchase program provides market liquidity on international financial markets. It can thus stimulate consumption and investment domestically and internationally. This leads to a boost in the domestic and foreign economy affecting output and prices.

The signaling channel affects the economy through central bank communication. A central bank can communicate via statements, speeches, policy announcements or reports on the economic development. By releasing information about central bank expectations, it provides new information to market participants, and may influence expectations. In particular, investors might adjust their expectations for future short-term interest rates in response to an announcement of a large asset purchase program, which directly affects long-term interest rates. Similarly, weaker growth forecasts announced by the central bank causes a change in expectations of investors and leads to lower long-term yields and lower asset prices.

3.4 Methodology

3.4.1 Three-Country Structural VAR Model

We analyze international monetary policy spillovers using a three-country vector autoregressive model with variables for the US, UK, and EA.[6] The VAR model for one country i with $i = 1, 2, 3$, G_i country-specific variables and $G = G_{US} + G_{UK} + G_{EA}$ variables of all countries for $t = 1, ..., T$ is given by[7]

$$y_{it} = c_i + \phi_{i1} y_{t-1} + \phi_{i2} y_{t-2} + ... + \phi_{ip} y_{t-p} + u_{it},\ u_{it} \sim \mathcal{N}(0, \Sigma_{ii}). \tag{3.1}$$

The c_i is a country-specific intercept term and ϕ_{iP} are the $[G_i \times G]$-coefficient matrices for lags $P = 1, ..., p$. The G_i country-specific variables for one country i are given in y_{it} which is a $[G_i \times 1]$-vector. The $y_{t-P} = \left[y'_{1t-P}, y'_{2t-P}, y'_{3t-P}\right]'$ contains the lagged variables of all countries for lags $P = 1, ..., p$. Thus, the model allows for dynamic interdependencies across economies. That means, lagged endogenous variables of foreign

[6]Note that we use the term "country" also to describe the euro area.

[7]Note that we will later relax the assumption of a common sample size T for all models.

countries can impact variables of the domestic country. The country-specific intercept term and coefficient matrices enable to model country heterogeneities in the three-country VAR model. The u_{it} is normally distributed with covariance matrices across countries Σ_{ij} for $i \neq j$.

Including the three countries in one model, the large reduced form three-country VAR model can be written as

$$y_t = \Phi x_{t-1} + u_t,\; u_t \sim \mathcal{N}(0, \Sigma), \tag{3.2}$$

where Φ is a $[G \times Gp + 1]$-coefficient matrix and $x_{t-1} = [1, y_{t-1}, ..., y_{t-p}]'$ is the matrix of lagged endogenous variables.[8] The covariance matrix Σ is of dimension $[G \times G]$. The covariance matrix is unrestricted. Thus, the model allows for correlation among all country and variable combinations.

The three-country structural vector autoregressive (SVAR) model in compact form is given by

$$A_0 y_t = A x_{t-1} + e_t, \tag{3.3}$$

where A_0 is the invertible matrix of contemporaneous dependencies and $A = \left[c', A_1', \ldots, A_p'\right]'$ contains a constant and lagged dependencies. The $y_t = [y_{1t}', y_{2t}', y_{3t}']'$ contains the variables of all countries. The $x_{t-1} = [1, y_{t-1}, ..., y_{t-p}]'$ is the matrix of lagged endogenous variables. The vector of structural shocks, e_t, is of dimension $[G \times 1]$.

The structural form parameters are linked to the reduced form parameters through $\Sigma = (A_0' A_0)^{-1}$ and thus $\Phi = A_0^{-1} A$. Let Σ_{tr} be the lower triagonal Cholesky component of Σ and Ω any $[G \times G]$ orthogonal rotation matrix. The vector of reduced form errors, u_t, is linked to the structural errors, e_t, by:

$$u_t = A_0^{-1} e_t = \Sigma_{tr} \Omega e_t. \tag{3.4}$$

Two different orthogonal matrices Ω and $\tilde{\Omega}$ can give different structural parameters without affecting the likelihood of the model which only depends on the reduced-form errors. Thus, to identify the structural parameters additional restrictions are needed. The identifying restrictions are set on Ω by choice of a suitable prior distribution. As we use a Bayesian proxy SVAR model, the identifying restrictions are additional information coming from including the proxy data. That is, the proxy informs the choice of Ω, and the prior on Ω is updated based on the proxy data.

[8] The exact estimation of parameters in Φ under a Bayesian selection prior is explained in section 3.4.4.

3.4.2 Identification of Monetary Policy Shocks

We use a high frequency approach to identify monetary policy shocks in a structural VAR model. The general idea of this approach is to use an exogenous series which is highly correlated to the structural shock of interest and unrelated to all other structural shocks. Loosely speaking, the structural shock of interest can be identified by employing the exogenous series as an instrument. Faust et al. (2003) as well as Faust et al. (2004) introduce to use high frequency data as measure for monetary policy shocks in SVAR models. The authors follow the idea of the event study literature to use high frequency financial data to identify monetary policy shocks (e.g., Kuttner, 2001; Gürkaynak et al., 2005). Faust et al. (2003) identify a monetary policy shock by matching the impulse responses of exchange rates and interest rates to the shock in an open economy VAR model to the changes of the high frequency financial data around monetary policy announcements of the Fed. Faust et al. (2004) use the matching requirement for the response of interest rates in a VAR model to those from federal funds futures data to the monetary policy shock. Recently, Stock and Watson (2012) and Mertens and Ravn (2013) use external instruments for structural identification in macroeconomic dynamic time series models. Stock and Watson (2012) identify six structural shocks in a high-dimensional dynamic factor model using constructed shocks given in the literature as external instruments. Mertens and Ravn (2013) identify tax shocks in a VAR model with the help of narratively identified changes in tax liabilities. Gertler and Karadi (2015) use a proxy series based on high frequency data to identify a monetary policy shock in a structural VAR model.

The monetary policy surprise series are constructed as the price changes of future government bond contracts around monetary policy announcements. It is based on the assumption that the only additional information affecting interest rate futures (i.e., financial market expectations on future interest rates) in a short window around central bank communication events is the central bank communication itself. Under this assumption, high-frequency information on futures can be used to create a proxy for the size and direction of monetary policy shocks. Importantly, the changes in future prices reflect the adjustment of the expectations of market participants regarding (i) the future movement of the target rate, (ii) forward guidance regarding longer-term plans of the central bank and (iii) information on central bank forecasts of economic development.

Let $m_{i,t}$ be the proxy constructed from the changes of high frequency government bond future contracts of one country i around policy announcement days of the country's central bank. Partition the $[G \times 1]$-vector of structural shocks of the SVAR model given in equation (3.3), e_t, into $e_t = [e_{MPi,t}, e'_{2,t}]'$. Let $e_{MPi,t}$ be the shock of interest.

The $e_{MPi,t}$ is the structural monetary policy shock of country i in the VAR model including variables from all three economies. The $[(G-1)\times 1]$-vector $e_{2,t}$ contains all other $G-1$ structural shocks. The proxy $m_{i,t}$ needs to fulfill the following two conditions to identify the structural monetary policy shock of one specific country, $e_{MPi,t}$:

$$E(m_{i,t}e_{MPi,t}) \neq 0$$
$$E(m_{i,t}e'_{2,t}) = \mathbf{0}'.$$

The first condition implies that the proxy needs to be correlated to the structural monetary policy shock of interest. That means, the proxy can be a noisy measure of the structural shock but needs to carry some information about it. The second condition means that the proxy needs to be uncorrelated to all remaining structural shocks of the VAR model, including foreign monetary policy shocks. By using high frequency data to calculate the price changes, the assumption of no correlation can be justified. Importantly, its validity with respect to actually identified shocks can also be tested. In our case, it suffices to test the correlation of proxies. Thus, the proxy $m_{i,t}$ is informative about the monetary policy shock of one country and is orthogonal to all other structural shocks in the VAR model, $e_{2,t}$.

We aim at identifying one by one three monetary policy shocks: for the US, UK, and EA. The identification for each shock is based on a proxy series from the literature, which is constructed from respective government bond future contracts and monetary policy actions of the Fed, BoE or ECB.

Under the validity of the second assumption above, it is possible to focus in one VAR model only on the identification of one monetary policy shock of a specific country. The reason for this is that the external series are orthogonal to each other. Each series thus does indeed only contain information on one monetary policy shock. An alternative way is to jointly identify three monetary policy shocks in one model using three proxies, which is computationally less efficient.

3.4.3 Bayesian Proxy Three-Country Structural VAR Model

We use a Bayesian proxy three-country structural VAR model following Caldara and Herbst (2016). The model leans on Stock and Watson (2012) and Mertens and Ravn (2013) using a Bayesian implementation. We estimate a three-country VAR model and identify the three shocks, $e_{MP,US,t}, e_{MP,UK,t}$, and $e_{MP,EA,t}$, individually in succession. We successively augment the VAR, first, with a proxy series constructed for the US to identify the US monetary policy shock, second, with a proxy for the UK for the identification of the UK monetary policy shock and, third, with a proxy for the EA to identify the EA monetary policy shock. A proxy constructed around announcement days of one of the three central banks is linked to the unobserved monetary policy shock of the same country i for $t = 1, ..., T$ by[9]

$$m_{i,t} = b_i e_{MPi,t} + \sigma_{v,i} v_{i,t},\ v_{i,t} \sim \mathcal{N}(0,1) \text{ and } v_{i,t} \perp e_t. \tag{3.5}$$

Equation (3.5) captures the two conditions for the proxies. First, if the coefficient b_i is nonzero, the proxy is informative about the structural shock, $e_{MPi,t}$. The ratio of $b_i/\sigma_{v,i}$ can be seen as a measure of the informativeness of the proxy for the structural shock. A large ratio of b_i to $\sigma_{v,i}$ means that the proxy is informative about the structural shock. Weak identification is present, if the ratio is small and, thus, if the proxy is less informative about the monetary policy shock. If $b_i = 0$, the proxy is not able to identify the monetary policy shock. Second, the proxy is orthogonal to all other structural shocks of the system, $v_{i,t} \perp e_t$.

An important feature of the identification via high frequency data in the Bayesian proxy SVAR model is that the proxy can be a noisy measure of the structural shock. Only the assumptions stated beforehand of non-zero correlation between the proxy and the identified shock and no correlation with all other shocks are needed. The reliability of the proxy can then be investigated by considering the posterior distributions of b_i and $\sigma_{v,i}$.

For simplicity, the variables in the SVAR model are ordered such that the structural shock of interest is ordered first. The monetary policy indicator variable is the first variable in the system. Thus, the first equation of the SVAR model given in (3.3) is the monetary policy equation:

$$A_{0,[1\bullet]} y_t = A_{[1\bullet]} x_{t-1} + e_{MPi,t} \tag{3.6}$$

where $A_{0,[1\bullet]}$ and $A_{[1\bullet]}$ denote the first row of matrix A_0 and A, respectively. Combining equations (3.5) and (3.6) gives the following relation of the monetary policy equation of the SVAR model and the proxy equation:

$$\begin{aligned} m_{i,t} &= b_i(A_{0,[1\bullet]} y_t - A_{[1\bullet]} x_{t-1}) + \sigma_{v,i} v_{i,t} \\ -\frac{b_i}{\sigma_{v,i}} A_{0,[1\bullet]} y_t + \frac{m_{i,t}}{\sigma_{v,i}} &= -\frac{b_i}{\sigma_{v,i}} A_{[1\bullet]} x_{t-1} + v_{i,t}. \end{aligned} \tag{3.7}$$

[9]Note that we will relax the assumption that the sample size, T, is equal for the proxy data and the VAR data.

The proxy three-country SVAR model augments the SVAR model with the proxy equation. Combining (3.3) and (3.7) gives:

$$\begin{bmatrix} A_0 & 0 \\ -\frac{b_i}{\sigma_{v,i}} A_{0,[1\bullet]} & \frac{1}{\sigma_{v,i}} \end{bmatrix} \begin{bmatrix} y_t \\ m_{i,t} \end{bmatrix} = \begin{bmatrix} A \\ -\frac{b_i}{\sigma_{v,i}} A_{[1\bullet]} \end{bmatrix} x_{t-1} + \begin{bmatrix} e_t \\ v_{i,t} \end{bmatrix}. \tag{3.8}$$

The structural parameters A_0 and A are a function of the reduced from parameters Φ, the reduced form covariance matrix Σ, and a rotation matrix Ω. That is, $A_0^{-1} = \Sigma_{tr}\Omega$ with Σ_{tr} denoting the lower triangular Cholesky decomposition of Σ, and $A = \Sigma_{tr}\Omega\Phi$.

So far, an equal sample size, T, for the proxy data and VAR data as well as along different model specifications is assumed. It will get clear from the following expression of the joint likelihood given in (3.9), that it is possible to have different sample sizes for the proxy data and the VAR data. Furthermore, the beginning and end dates of the proxy and VAR data can vary across different model specifications depending on the variables (and countries) included. Let T_y denote the sample size of the VAR data, T_m the sample size of the proxy data, and t_{1m} the beginning of the proxy data. Let $y_{1:T_y}$ denote the data of the variables included in the VAR model. The sample size T_y varies across model specifications. The longest possible sample size is chosen for each model. Let $m_{i,t_{1m}:T_m}$ denote the proxy data for country i. The beginning of the proxy data, t_{1m}, can be later than the beginning of the VAR data. The beginning, t_{1m}, and end of the proxy data, T_m, varies across the proxies for the US, UK, and EA.[10] The exact sample sizes are given in section 3.5 (see for a summary table 3.1 in section 3.5.2).

The orthogonal matrix Ω does not depend on the VAR model data $y_{1:T_y}$ but is informed by the exogenous proxy data, $m_{i,t_{1m}:T_m}$. That is why, following Caldara and Herbst (2016), the joint likelihood of the Bayesian proxy three-country SVAR model can be expressed as

$$p(y_{1:T_y}, m_{i,t_{1m}:T_m}|\Phi, \Sigma, \Omega, b_i, \sigma_{v,i}) = p(y_{1:T_y}|\Phi, \Sigma) p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi, \Sigma, \Omega, b_i, \sigma_{v,i}). \tag{3.9}$$

The likelihood function is split into the first part which is a standard likelihood function for the reduced form VAR model parameters and the second part which gives the conditional likelihood of the proxy $m_{i,t_{1m}:T_m}$ given the data $y_{1:T_y}$, reduced form coefficients and the parameters of the proxy equation. The conditional likelihood $p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi, \Sigma, \Omega, b_i, \sigma_{v,i})$ is normally distributed with mean $\mu_{m_{i,t_{1m}:T_m}|y_{1:T_y}}$ and

[10] As the sample sizes for the VAR data and proxy data vary across models depending on which variables and countries are included, the exact notation would require an additional model specification index for the sample sizes. For a better clarity this index is neglected and the exact sample sizes are given in the model specification table 3.1 and are clearly marked below each graph in the result section.

variance $V_{m_{i,t_{1m}:T_m}|y_{1:T_y}}$ whereby

$$\begin{aligned}\mu_{m_{i,t_{1m}:T_m}|y_{1:T_y}} &= b_i\Omega'_{[\bullet 1]}\Sigma_{tr}^{-1}u_t = b_i e_{MPi,t} \\ V_{m_{i,t_{1m}:T_m}|y_{1:T_y}} &= \sigma_{v,i},\end{aligned} \tag{3.10}$$

where $\Omega_{[\bullet 1]}$ denotes the first column of Ω.

We estimate the model using a Metropolis-within-Gibbs algorithm. The split of the joint likelihood into the standard likelihood of $y_{1:T_y}$ and the conditional likelihood of $m_{i,t_{1m}:T_m}$ enables the separate estimation of the reduced form parameters and the structural parameters. However, the reduced form parameters are updated in light of the proxy. That is, draws from the posteriors $p(\Phi, \Sigma|y_{1:T_y})$ are accepted based on the evaluation at the conditional likelihood of $m_{i,t_{1m}:T_m}$. The algorithm combines a Gibbs sampler for coefficient draws with a Metropolis acceptance step, where new draws are accepted based on their effect on the likelihood of the proxy $p(m_{i,t_{1m}:T_m}|\cdot)$. Thus, we do not have standard reduced form posteriors but posterior distributions updated based on the structural identification. The parameter draws of Φ and Σ are not only conditional on the data but on the proxy data and parameters coming from the structural form, Ω, b_i, and $\sigma_{v,i}$. Hence, our proxies inform the estimation of the reduced and structural form parameters. The exact updating of reduced form parameters is given in 3.B.

The Bayesian proxy three-country SVAR model has several advantages. First, the influence of the proxies on reduced form parameters implies that more weight is put on those parameters that are consistent with the structural shocks. The strength of this "shift" depends on the informativeness of the proxy, which ensures that a clear identification of structural errors is accompanied by accepting reduced form coefficients based on the proxy data.

Second, the separate estimation of reduced form parameters and shock identification in one draw allows to have different numbers of observations T_y and T_m. Similar to the majority of the literature on proxy VAR models, we only require that proxy observations are a subset of observations of the data of endogenous variables. This is useful since in most application scenarios the high frequency data for the proxies are only available for a shorter period starting at a later date than the data for the macroeconomic variables ($T_y > T_m$). Still, for the variables included in the VAR model data starting earlier than the first available proxy data can be used for the estimation of the reduced form parameters. The matrices Φ and Σ are drawn based on longer time series, while the update step of the reduced form parameters and the draws of Ω, b_i, and $\sigma_{v,i}$ can use a shorter sample determined by the availability of the proxy data.

Third, the Bayesian approach simplifies the dealing with the weak identification problem and inference by adding subjective information. Weak identification leads to nearly flat likelihoods and thus a strong dependence of the posterior on the prior distribution. However, it still allows to sample from the posterior distribution. The specification of the prior distributions for the parameters of the proxy equation (3.5) adds additional (subjective) information to the model impacting strongly the posterior distributions.

Fourth, especially in the panel framework a large number of parameters have to be estimated. Since the reduced form can be estimated in a separate step, we can set a specific suitable prior distribution for our three-country model which remains independent of the proxy. Through shrinkage this prior allows us to estimate the large number of parameters assuming a panel structure.

3.4.4 Prior and Posterior Distributions

We use a Bayesian shrinkage approach for the estimation of reduced form parameters. Our Bayesian selection prior assumes a specific panel structure. It is thus an extension of the selection prior of Korobilis (2013). To ensure the estimation feasibility of our model the prior on reduced form coefficients reduces the dimension of the VAR model in two ways. First, the prior introduces Bayesian shrinkage. We take the commonly used Minnesota prior for VAR models. In the equation of variable y_i, this prior introduces a higher shrinkage on lags of variables y_j, $j \neq i$, and higher lags in general. This shrinkage is done by an increasingly tighter variance in the prior specification around a prior mean of zero. Second, the prior assumes a panel structure to set additional zero restrictions. The prior builds on the hypothesis that domestic variables are endogenously driven by all of their own lags, but not by all lags of all foreign variables. That is, some foreign lagged variables are excluded from the system. The exclusion is determined by a data based search for zero restrictions on the lagged coefficients of foreign variables while domestic lags are left unrestricted. The exclusion of coefficients facilitates estimation of the model even in cases of a large number of parameters to estimate.

The estimation procedure we are using follows the approaches to estimate panel vector autoregressive models. PVAR models combine several macroeconomic variables for multiple countries in a single model and therefore account for international spillovers. The characteristics of a PVAR model are that it captures dynamic and static interdependencies as well as heterogeneities across countries. That is, PVAR models include lagged variables of all countries in each equation (dynamic interdependencies), allow for correlation between error terms of all country and variable combinations (static interdependencies), and have country-specific intercepts and coefficient matrices (cross-country heterogeneities) (e.g., Canova and Ciccarelli, 2013). Our three-country VAR model is a PVAR model focusing on the relation between the US, UK, and EA. The estimation strategies for PVAR models use the panel structure of the model to set restrictions or reduce the parameters space. As yet, the literature on estimating PVAR models follows mainly three approaches: setting restrictions in line with theoretical arguments, using a Bayesian factor approach proposed by Canova and Ciccarelli (2004, 2009) or using Bayesian selection priors as done by Koop and Korobilis (2015b).

The dynamic interdependency restrictions we are searching for in our model are a typical characteristic of PVAR models. Searching for no dynamic interdependencies in PVAR models is motivated by the findings of Koop and Korobilis (2015b), Korobilis (2016), and Schnücker (2016). The authors provide evidence that the estimation can be improved by reducing mean squared estimation errors if the panel dimension of the data is accounted for in the specification of the prior. We allow for a flexible panel-structure since a zero restriction can be set on each coefficient separately not assuming that all variables of one country are restricted in the same way.

The idea of the selection prior is to select the non-zero parameters in the model depending on an indicator. We use a binary selection indicator γ for the inclusion decision of each coefficient in the model. If γ is set to zero, the coefficient is set to zero. If γ is one, the coefficient is freely estimated. The binary indicators on each variable follow a Bernoulli distribution with a posterior probability that scales a prior probability of inclusion by the likelihood gains in case of inclusion. Hence, the exclusion of variables is data-driven.

The zero restrictions of this model thus imply a spike-and-slab mixture distribution on reduced form coefficients, that is, a distribution with a point mass at zero (the spike) and the posterior distribution in case of $\gamma = 1$ (the slab). The zero restrictions in the model are introduced by assuming a mixture distributions on the reduced form coefficients. The lagged coefficients of foreign variables are either set to zero or drawn from a posterior distribution based on a Minnesota prior. The lagged coefficients of domestic variables are only based on the Minnesota prior (i.e., these coefficients are always included). The intercept term is also left unrestricted. Let ϕ be one arbitrary element of Φ, the $[G \times Gp + 1]$-coefficient matrix of the reduced form VAR model, and let γ be the corresponding binary selection indicator. Then the reduced form coefficient ϕ is drawn from the posterior distribution in the following way

$$
\begin{aligned}
&\phi = 0 && \text{if } \gamma = 0 \text{ for } \textit{foreign lag} \\
&\phi \sim p(\Phi | y_{1:T_y}, \Sigma) && \text{if } \gamma = 1 \text{ for } \textit{foreign lag} \\
&\phi \sim p(\Phi | y_{1:T_y}, \Sigma) && \text{for } \textit{domestic lag and constant.}
\end{aligned}
$$

If $\gamma = 1$ for foreign lags or ϕ is a domestic lag coefficient or a constant, ϕ is drawn from the posterior distribution $p(\Phi|y_{1:T_y}, \Sigma)$. If $\gamma = 0$ for foreign lags, ϕ is set to zero.[11]

As Caldara and Herbst (2016) we assume a mixture proposal distribution as the prior for the reduced form covariance matrix Σ. As explained above, we give the proxy the possibility to inform the posterior distribution of reduced form coefficients (in the form of a Metropolis acceptance step). That is, we allow for the possibility that the posterior of Σ given $y_{1:T_y}$ and $m_{i,t_{1m}:T_m}$ can be different from the posterior given $y_{1:T_y}$. This implies that drawing a candidate Σ from the latter posterior distribution may be inefficient. To improve on this, Caldara and Herbst (2016) propose the mixture in order to put additional weight on draws of the variance-covariance matrix that fit well to the proxy. Specifically, we draw Σ from a weighted average of two inverse Wishart distributions. The first distribution or, equivalently, its scaling matrix and degrees of freedom is determined by the Minnesota prior and the data and denoted by $p(\Sigma|y_{1:T_y})$. The second inverse Wishart distribution $\mathcal{IW}(\Sigma^{old}, d)$, uses the last draw Σ^{old} from the Gibbs sampler as scaling matrix, and prespecified degrees of freedom d. Let ρ be the weight of the posterior inverse Wishart distribution determined by the data, then the mixture proposal distribution for Σ is

$$p(\Sigma|y_{1:T_y}, \Sigma^{old}, d) = \rho p(\Sigma|y_{1:T_y}) + (1-\rho)\mathcal{IW}(\Sigma^{old}, d). \tag{3.11}$$

A $\rho < 1$ gives some weight on a distribution based on previous draws (instead of the data) and thus includes a random walk like behavior. That is, $\rho < 1$ gives the algorithm an additional opportunity beyond simple rejection of new draws to keep candidate draws for the variance-covariance matrix in regions that fit well to the proxy. Other than in the case of the selection indicator for the reduced form coefficients, the parameter ρ is not drawn from a distribution, but is set at a fixed value.

Further, we set the following prior distributions for the proxy equation. We take a normal prior for b_i, $b_i \sim \mathcal{N}(b_{0,i}, V_{b,i})$. For the standard deviation of the measurement error, $\sigma_{v,i}$, we use two different prior specifications: a so called high relevance prior and an inverse Gamma prior. The high relevance prior, following Caldara and Herbst (2016), states the subjective assumption that $\sigma_{v,i}$ is equal to a specific proportion, α, of the standard deviation of the proxy data, $\sigma_{m_{i,t_{1m}:T_m}}$, with probability 1. That is, the

[11] A detailed description of the implementation of such a selection prior can be found in Korobilis (2013). The prior is closely related to the selection prior proposed in chapter 1, but does not search for homogeneity and static interdependency restrictions. Moreover, the spike and slap prior used in this chapter sets actual zeros while the selection prior of chapter 1 shrinks coefficients to zero. By setting zeros on foreign lagged coefficients the prior used here is closely related to the frequentist lasso approach introduced in chapter 2.

high relevance prior is given by

$$\sigma_{v,i} = \alpha \times \sigma_{m_{i,t_{1m}:T_m}}.$$

The α is set to 0.2 or 0.4. The detailed prior choices are given in 3.D. This prior states a tight relationship between the proxy and the SVAR model. The posterior equals the prior as the prior is not updated by the data. We also try a less informative prior and use an inverse Gamma distribution for $\sigma_{v,i}$, $\sigma_{v,i} \sim \mathcal{IG}(s_i, S_{v,i})$. The parameter s_i is the degrees of freedom parameter and $S_{v,i}$ the centering coefficient.[12]

3.5 Data

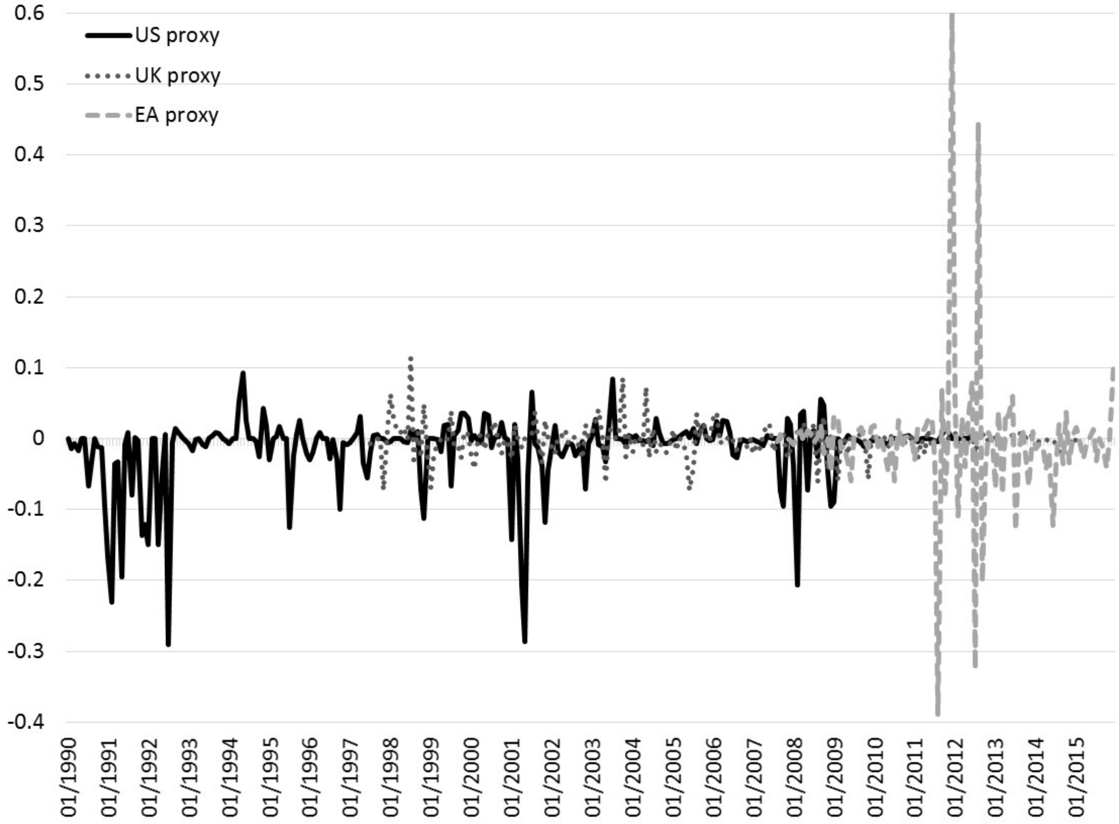
3.5.1 Proxy Series

For the identification with external instruments we use for the US the shock series constructed by Gertler and Karadi (2015). The proxy series for the UK is taken from Gerko and Rey (2017) and for the EA from Rogers et al. (2014, 2017). The proxy series are plotted in figure 3.2. They cover different time episodes. All three proxies capture the periods of conventional and unconventional monetary policy.

Gertler and Karadi (2015) construct the proxy series with changes in the three month ahead federal funds futures on FOMC dates between January 1990 and June 2012. They measure the price differences in a 30 minutes window around policy announcements, which carry the overwhelming amount of new information on monetary policy in the US, as argued already above. However, policy announcements also include information on macroeconomic expectations in the Fed. Thus, policy announcements provide two dimensions of information to market participants, which materializes in a mixture of monetary policy shocks and what has been called "central bank information shocks" in some of the literature (Nakamura and Steinsson, ming; Jarociński and Karadi, 2017). In our analysis, as in the large majority of the literature, we are not able to disentangle these two shocks. However, given the shock series in Figure 3.2, we can at least hypothesize on the relative strength of the latter shock: the Fed increased the amount of information contained in the policy announcements (*forward guidance*) over time, taking a large step forward during the recent period of interest rates at the zero lower bound. During the same period, the shocks should be largely driven by central bank information, as the zero lower bound was not expected to end soon and thus effectively anchoring futures at zero as well. As can be seen, these shocks are quite small during this period, compared to the larger shocks earlier in the sample.

[12] The posterior distributions are given in 3.A.

Figure 3.2: Monetary policy surprise series



Note: US monetary policy surprise series of Gertler and Karadi (2015), UK series of Gerko and Rey (2017), EA series of Rogers et al. (2017).

Moreover, we can savely argue that central bank information shocks are most likely smaller in the earlier part of the sample than during the period at the zero lower bound, as a lower amount of information should also lead to smaller shocks. Another critique raised by Miranda-Agrippino (2016) states that the proxy series may be predictable both by publicly available financial market variables and private central bank forecasts. Although the differences between "raw" proxies and their unpredictable component are not large, the former may result in sometimes counterintuitive impulse-response functions. However, she notes that this problem can be accounted for in two ways. First, we could use the unexplained part of the proxy as an instrument instead of the "raw" proxy. Second, we could account for the publicly available information in the VAR model directly by including a risk variable like the excess bond premium in the US or comparable measures in the UK and Eurozone, see also Caldara and Herbst (2016) for a longer discussion of this issue.

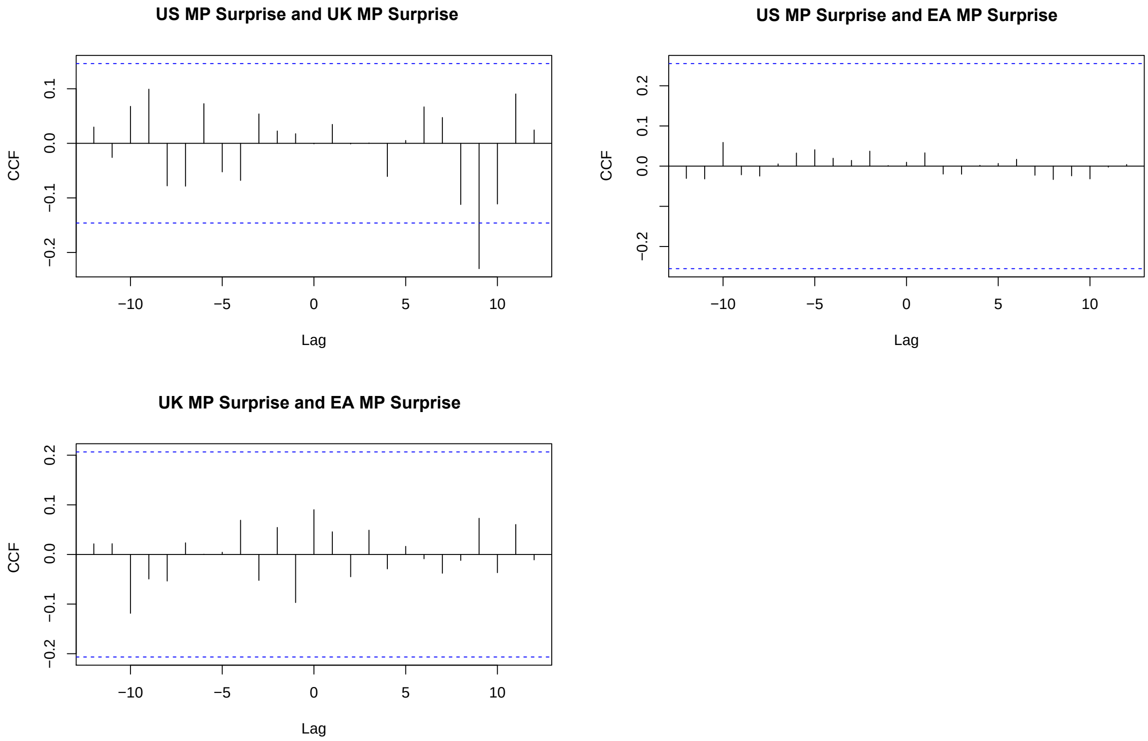
Gerko and Rey (2017) use three months short-sterling futures for the UK between January 2000 and January 2015. They construct the price changes in a window ten minutes before to twenty minutes after the monetary policy announcement. Other than in the US, the timing structure of monetary policy actions of the BoE allows to disentangle the effect of target rate announcements and central bank information about the future policy actions or future economic developments. Since the BoE releases minutes of meetings and inflation reports on different dates than short press announcements only containing the actual interest rate decision, the proxies can be constructed separately for the two policy actions. As Gerko and Rey (2017) we use the proxy constructed around minutes and inflation reports. Rate announcements release information about changes in the policy rate and asset purchases. During the recent period of constant low interest rates the rate announcements carry little information about the monetary policy. The proxy constructed around minutes and inflation reports, however, is informative about monetary policy actions since it reveals information about forward guidance. In contrast, since the federal funds futures of the US capture information about rate changes but also due to the timing information about the central bank's view of the economy, the constructed proxy aggregates information about rate announcements and forward guidance. Thus, the proxy is also informative for the zero-lower bound period.[13]

For the EA instrument, Rogers et al. (2017) use the spread between Italian and German 10-year government bond yields as the external instrument following Rogers et al. (2014). Rogers et al. (2014) argue that taking the spread between Italian and German yields is done to account for the fact that the unconventional monetary policy of the ECB was aimed at reducing the intra-European yields rather than affecting German yields. The changes are constructed in a fifteen minutes before and one hour forty-five minutes after window around policy announcements for August 2007 to December 2015. In months where no policy announcement takes place the price change is set to zero, while for more than one announcement per month the price changes are summed up.

By analyzing both conventional and unconventional monetary policy in a linear model we assume no structural changes in the monetary policy transmission. As argued above, while the changes of future government bond rates are mainly driven by rate announcements during conventional monetary policy periods, futures react mainly to forward guidance measures during unconventional monetary policy periods. Although the magnitude of the composition of the monetary policy shocks changes during the periods, the shocks always capture announcement effects due to new information on monetary policy and on macroeconomic expectations of the central banks. We fol-

[13]See also the discussion in Gerko and Rey (2017) about the information content of rate announcements and minutes and inflation reports.

Figure 3.3: Cross-correlations of monetary policy surprise series



Note: Cross-correlation functions for US monetary policy surprise series of Gertler and Karadi (2015), UK series of Gerko and Rey (2017), and EA series of Rogers et al. (2017). The lag values on the x-axis show the cross-correlation between one proxy at t and the second proxy at $t \pm lag$. The dotted lines depict the critical values at the 5% level.

low recent papers such as Gertler and Karadi (2015), Miranda-Agrippino (2016), and Cesa-Bianchi et al. (2016) by analyzing conventional and conventional monetary policy jointly in a linear model.

The exogenous variables we use as proxies of our three structural shocks should ensure that the monetary policy shocks in the US, UK, and EA are orthogonal to each other, both contemporaneously and intertemporally. Due to the linear and univariate relationship between proxy and shock, a necessary although not sufficient condition for this is that these properties also hold for the proxies. The contemporaneous correlations between the proxy series is negligible, for the proxies of US and UK -0.0009, US and EA 0.0097, EA to UK 0.0902. Figure 3.3 shows the intertemporal correlations between the proxy series. The graph on the upper left depicts the cross-correlation function for twelve leads and lags of the US and UK monetary policy proxies. The cross-correlation function, $CCF(m_{US}, m_{UK})$, gives the correlation between $m_{US,t+k}$ and $m_{UK,t}$ for $k = -12, ..., 0, ...12$. The value of the cross-correlation function at lag zero shows the contemporaneous correlation. The graph on the upper right shows the

Table 3.1: Model specifications: Individual country VAR models

	VAR data	proxy data	variables included
Individual country VAR models			
(1)	Jan 1973 to June 2017	Jan 1990 to June 2012	US 1-Y rate, US CPI, US IP, US UN, US EBP, US Dow Jones Index, US 10-Y rate, US prime rate banks
(2)	Jan 1982 to Mar 2017	July 1997 to Jan 2015	UK 5-Y rate, UK CPI, UK IP, UK UN, UK mortgage spread, UK FT index, UK 10-Y rate, UK prime lending rate
(3)	Jan 2000 to Oct 2017	Oct 2007 to Dec 2015	Italian 5-Y rate, EA CPI, EA IP, EA UN, EA mortgage spread, EA Euro share index, EA 10-Y yield, EA loans to non financial corporations
Individual country VAR models with exchange rates			
(4)	Jan 1973 to June 2017	Jan 1990 to June 2012	US 1-Y rate, US CPI, US IP, US UN, US EBP, US Dow Jones Index, US 10-Y rate, US prime rate banks, exchange rates US to UK and US to EU
(5)	Jan 1982 to Mar 2017	July 1997 to Jan 2015	UK 5-Y rate, UK CPI, UK IP, UK UN, UK mortgage spread, UK FT index, UK 10-Y rate, UK prime lending rate, exchange rates US to UK and EU to UK
(6)	Jan 2000 to Oct 2017	Oct 2007 to Dec 2015	Italian 5-Y rate, EA CPI, EA IP, EA UN, EA mortgage spread, EA Euro share index, EA 10-Y yield, EA loans to non financial corporations, exchange rates US to EU and EU to UK

cross-correlation function for the US and EA and the bottom graph for the UK and EA. All plots provide no evidence for cross-correlation among the proxy series, contemporaneously and for all leads and lags. One exception is the correlation between the US proxy at $t + 9$ and the UK proxy at t which lies outside the 95% confidence interval. However, given a total of 75 tested correlations, we should expect a larger number of tests which fail to reject. There would also be no economic explanation why there should be a negative correlation with a nine months lag.

3.5.2 Variables in the VAR Model

We empirically investigate monetary policy spillovers in nine different model specifications. A summary of the individual country VAR models and models including exchange rates is given in table 3.1 and of the three-country VAR models in table 3.2. The first three specifications are individual country models including only variables of one country. We include eight country-specific variables in the single-country VAR

Table 3.2: Model specifications: Three-country VAR models

	VAR data	proxy data	variables included
Three-country VAR models			
(7)	Jan 1991 to June 2017	Jan 1990 to June 2012	US 1-Y rate, US CPI, US IP, US EBP, US Dow Jones Index, US 10-Y rate, exchange rates US to UK and US to EU, UK CPI, UK IP, EA CPI, EA IP
(8)	Jan 1991 to Mar 2017	July 1997 to Jan 2015	UK 5-Y rate, UK CPI, UK IP, UK mortgage spread, UK FT index, UK 10-Y rate, exchange rates US to UK and EU to UK, US CPI, US IP, EA CPI, EA IP
(9)	Jan 2000 to Oct 2017	Oct 2007 to Dec 2015	Italian 5-Y rate, EA CPI, EA IP, EA mortgage spread, EA Euro share index, EA 10-Y yield, exchange rates US to EU and EU to UK, US CPI, US IP, UK CPI, UK IP

models: monetary policy indicator, log consumer price index, log industrial production, unemployment rate, credit spread, log stock price index, a long term interest rate, and a lending rate, all measured at monthly frequency. Model (1) includes the variables for the US. The model is augmented with the US proxy to identify the US monetary policy shock. The single-country VAR model for the US is estimated with data from January 1973 to June 2017 for the variables included in the VAR model and proxy data from January 1990 to June 2012. Model (2) is the individual country model for the UK including the UK variables and identifies the UK monetary policy shock. The single-country VAR model for the UK uses data from January 1982 to March 2017 and proxy data from July 1997 to January 2015. To analyze spillovers from the EA monetary policy shock to EA variables, model (3) is used in which EA variables are included and the EA monetary policy shock is identified. The EA model is based on data from January 2000 to October 2017 and proxy data from October 2007 to December 2015.

Models (4), (5), and (6) additionally include exchange rates. Model (4) adds to the variables included in model (1) exchange rates from the Euro and British pound to the US dollar and covers the same sample period as model (1). Model (5) extends model (2) by including exchange rates from the Euro and US dollar to the British pound for the same time period as in model (2). Model (5) includes exchange rates with respect to the Euro and the variables of model (3) for the same period as in model (3).

The three-country VAR models (7), (8), and (9) include variables for all countries jointly. Model (7) is augmented with the US proxy and thus the US monetary policy shock is identified. The model includes twelve variables. UK and EA CPI and IP are added to six US variables and the exchange rates with respect to the US dollar.

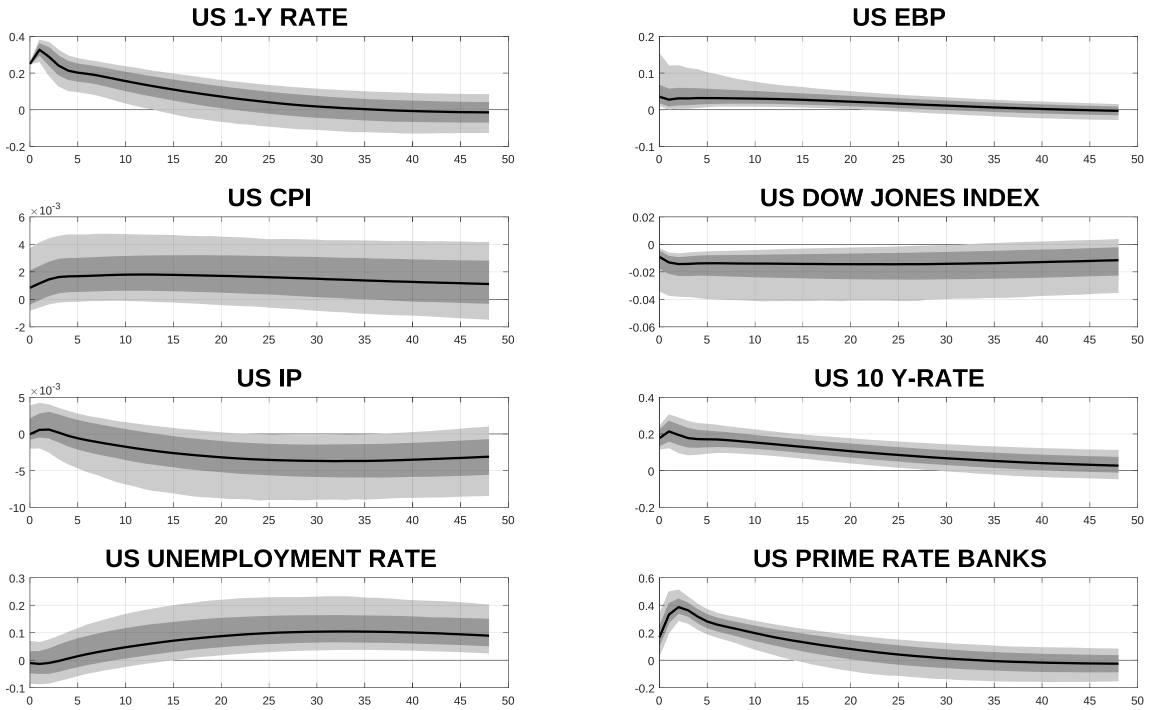
The data range from January 1991 to June 2017 (proxy data from January 1991 to June 2012). Model (8) is the UK version of model (7). That is, the UK monetary policy shock is identified with the UK proxy, six UK variables, the exchange rates to the British pound and US and EA CPI and IP are included. The data range from January 1991 to March 2017 (proxy from July 1997 to January 2015). Model (9) is the multi-country model where the EA monetary policy shock is identified. It adds to six EA variables the exchange rates to the Euro and US and UK CPI and IP. Data cover the period from January 2000 to October 2017 (proxy from October 2007 to December 2015). The data coverage for all models is determined by the availability of the data and the longest possible sample size is chosen.

Besides the log consumer price index (CPI) as a measure for prices and the industrial production (IP) variable as a measure of real activity, the unemployment rate (UN) is also included in the models. A credit spread variable is added to account for financial channels, as this improves the estimation of the model (Gertler and Karadi, 2015; Caldara and Herbst, 2016; Miranda-Agrippino, 2016). The credit spread variable is the excess bond premium for the US (US EBP), and mortgage spreads for the UK and EA (i.e., the differences between the mortgage rate and bank rate or three month rate, respectively Euribor). Stock prices capture financial spillover channels. The stock price index is the Dow Jones industrial share price index for the US (US DOW JONES INDEX), the Financial Times Stock Exchange index for the UK (UK L FT INDEX), and the Datastream Euro share price index for the EA (EA EURO SHARE INEX). Stock prices may increase due to the portfolio balance channel or be negatively affected by the signaling channel. Long term government bonds and the lending rate as a measure of financial conditions capture financial spillovers as well. The long term rates are the US 10-year treasury rate (US 10-Y RATE), the 10-year British government securities (UK 10-Y RATE), and the EA 10-year government bond yield (EA 10-YEAR YIELD). The lending rates are the US prime rate charged by banks (US PRIME RATE BANKS), the UK major banks prime lending rate (UK PRIME LENDING RATE), and the EA loans to non financial cooperations (EA LOANS NON FINL CORPS). We add exchange rates to capture the exchange rate channel for monetary policy transmission.

The monetary policy indicator is the 1-year treasury yield for the US (US 1-Y RATE), the 5-year yield from British government securities for the UK (UK 5-Y RATE), and the 5-year Italian government bond yield for the EA (ITALIAN 5-Y RATE).[14] By taking a long term government bond rate as the monetary policy indicator in the structural VAR model instead of short term interest rates such as the federal funds rate, the model

[14]Details on the data are given in 3.C.

Figure 3.4: Model (1): Responses of US variables to a contractionary US monetary policy shock



Note: US monetary policy surprise series of Gertler and Karadi (2015). Responses of US variables to a 25 basis point increase in the US treasury 1-year rate. A sign restriction is set on the response of the excess bond premium on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

can capture unconventional monetary policy action. Impulse responses are normalized to a 25 basis point increase in the monetary policy variable.

3.6 Results

3.6.1 Domestic Monetary Policy Transmission

Figure 3.4, 3.5, and 3.6 show the dynamic responses of domestic variables to a domestic contractionary monetary policy shock for the US, UK, and EA in the individual country models (1), (2), and (3). The impulse response functions are based on models with tight priors for the standard deviation of the measurement error in the proxy equation, $\sigma_{v,i}$. Results based on the less informative inverse Gamma prior are given in 3.F.2. All models include four lags and a constant.[15]

We find similar reactions of stock prices to domestic monetary policy shocks for the US, UK, and EA. A monetary tightening causes a drop of the country's stock price index. Furthermore, the monetary policy shocks of the US and UK increase the 10-year rate of the respective country. In general, monetary policy shocks identified around Fed and BoE announcements cause more pronounced responses of domestic variables than the shock by the ECB. This finding might be partly explained by the identification strategy and partly by analyzing the aggregated effects on the euro area and not on individual euro area countries.

The 25 basis point increase in the US 1-year treasury yield leads to a lagged increase in the unemployment rate as shown in figure 3.4 (model 1). CPI and industrial production do not react significantly. The monetary tightening causes a positive reaction of the excess bond premium. This is in line with the findings of Gertler and Karadi (2015) and Gerko and Rey (2017) who show an increase in the excess bond premium on impact and in line with the sign restriction set. The Dow Jones price index decreases on impact as expected for the signaling channel. The monetary tightening raises the ten year rate by around 20 basis points. In reaction to the monetary tightening lending rates increase comparable to the reaction of the 1-year rate impairing the financial conditions.

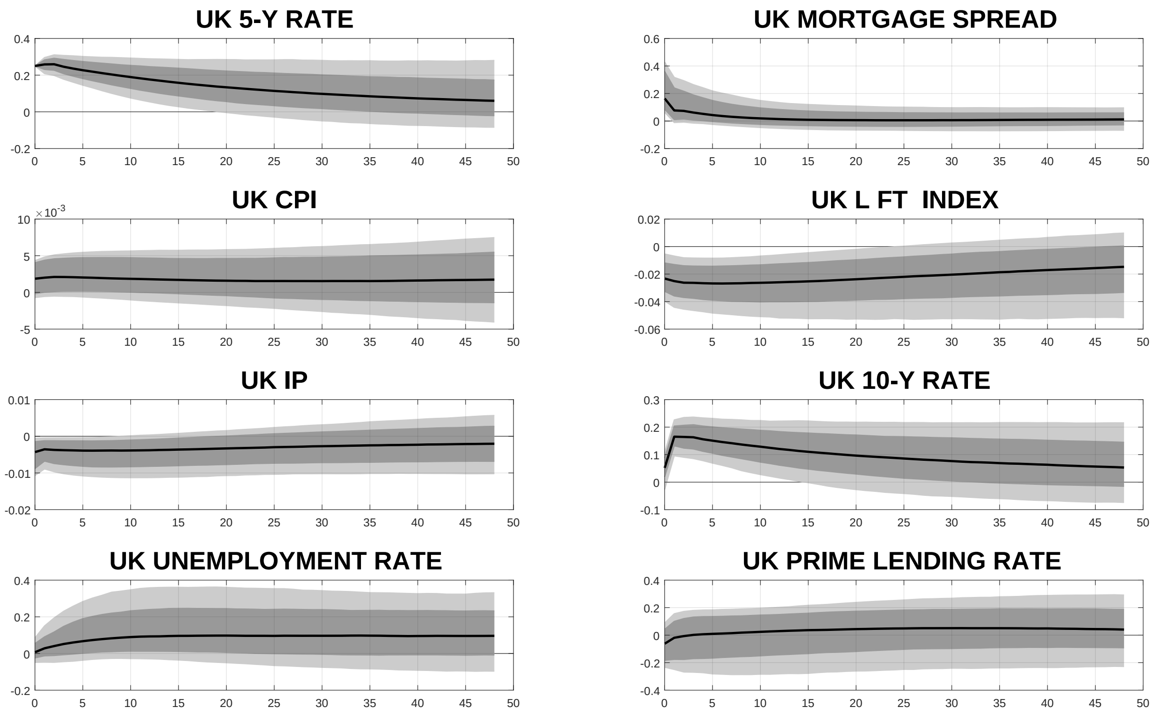
A monetary tightening in the UK has no clearly significant impact on CPI and on the unemployment rate and a modest negative impact on IP as shown in figure 3.5 (model 2).[16] The 25 basis point increase in the 5-year yield from British government securities leads to a modest increase of the mortgage spread on impact, which is in line with previous findings of Gerko and Rey (2017) and Cesa-Bianchi et al. (2016). As in the US, the stock price index declines significantly. The monetary tightening causes an increase of the 10-year rate with a peak impact of just below 20 basis points. The lending rate does not react significantly.

The US and UK model include an additional identifying restriction. A sign restriction is set on the responses of the US excess bond premium and the UK mortgage

[15] The Metropolis-within-Gibbs algorithm is based on 10,000 draws with a burn-in phase of 1000 draws.

[16] Figure 3.23 in 3.F.4 shows the results for the UK monetary policy shock using changes around rate announcements as the proxy series constructed by Gerko and Rey (2017) instead of the proxy constructed around minutes and inflation reports. The responses are not significant, in particular also the response of the 5-year rate. The findings support the view that changes around rate announcements carry only limited information about the monetary policy actions during the zero lower bound period. Figure 3.22 presents the results for the UK monetary policy shock of Rogers et al. (2017). The instrument is constructed as the changes in long gilt futures yield around meetings of the Monetary Policy Committee and asset purchase announcements. Inflation reports and the detailed minutes are rarely included. Thus, the results might also be explained by the construction of the proxy not carrying enough information regarding the monetary policy actions of the BoE.

Figure 3.5: Model (2): Responses of UK variables to a contractionary UK monetary policy shock

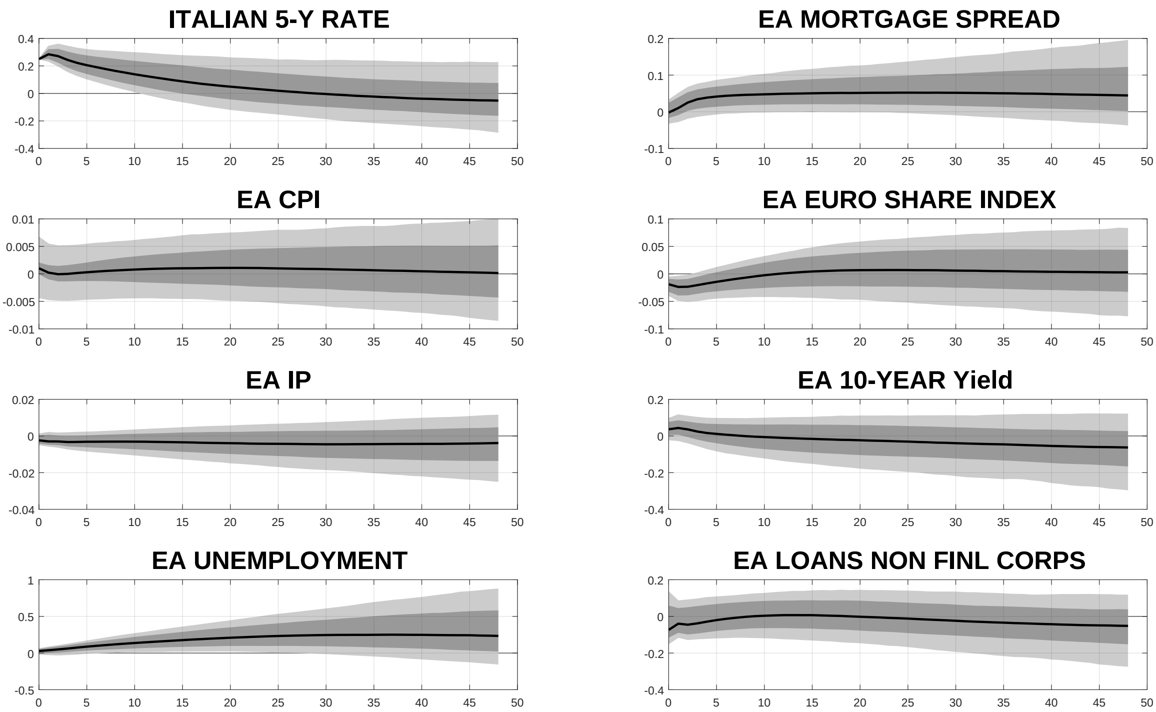


Note: UK monetary policy surprise series of Gerko and Rey (2017). Responses of UK variables to a 25 basis point increase in the 5-year yield from British government securities. A sign restriction is set on the response of the mortgage spread on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1982 to March 2017. Proxy data from July 1997 to Jan 2015.

spread to the monetary policy shock. The responses are restricted to be positive. The sign restriction is in line with the evidence of Gertler and Karadi (2015) and Gerko and Rey (2017) who find positive reactions of the excess bond premium and the mortgage spread, respectively. The impulse response functions without the sign restrictions are shown in 3.F.1. Both spreads respond negatively to the monetary policy shock causing a price puzzle in the case of the US and a positive reaction of UK industrial production and the unemployment rate.

A 25 basis points increase in the Italian 5-year rate causes a significant modest drop in the EA stock price index on impact as shown in figure 3.6 (model 3). The unemployment rate rises significantly after around five months while CPI and IP do not react significantly. The EA mortgage spread increases modestly with the raise getting barely significant after around seven months. The 10-year rate increases on impact,

Figure 3.6: Model (3): Responses of EA variables to a contractionary EA monetary policy shock

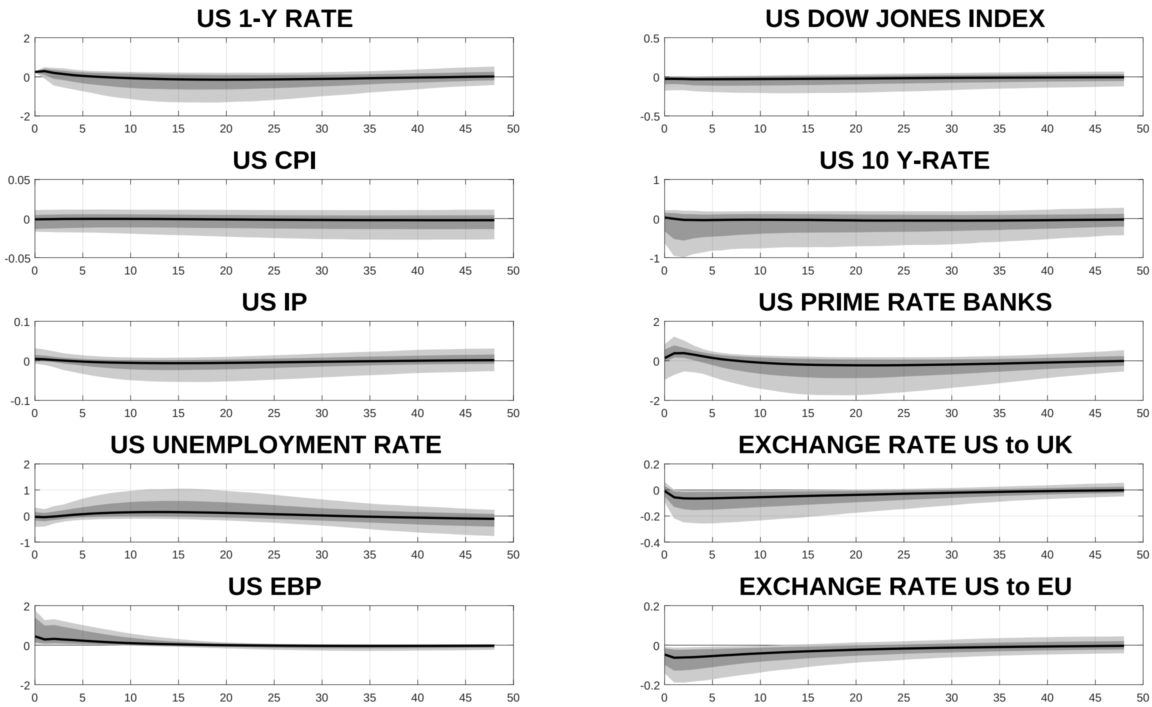


Note: EA monetary policy surprise series of Rogers et al. (2017). Responses of EA variables to a 25 basis point increase in the Italian gross yield of benchmark 5-year rate. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 2000 to Oct 2017. Proxy data from Oct 2007 to Dec 2015.

albeit insignificantly. Loans to non financial corporations also react insignificantly. The insignificant effects could be mainly driven by the identification strategy. An issue with the EA model is that a particularly tight prior has to be set such that the identified monetary policy shock leads to a response of the Italian 5-year rate without large credible sets. The proxy for the EA seems to carry less information to identify the monetary policy shock compared to the other two proxies and thus requires a tighter prior on the standard deviation of the measurement error in the instrument equation, which facilitates estimation of the linear relationship between proxy and the monetary policy shock.[17] Furthermore, it may be particularly difficult to find effects on the aggregated EA level as the single EA countries might react in various ways to a monetary

[17]The prior specifications are explained in 3.D. Additional results for the EA 2-year government bond rate as the policy indicator are in 3.F.3. The proxy seems to carry insufficient information to identify a monetary policy shock with the EA 2-year government bond rate as the policy indicator.

Figure 3.7: Model (4): Responses of exchange rates and US variables to a contractionary US monetary policy shock



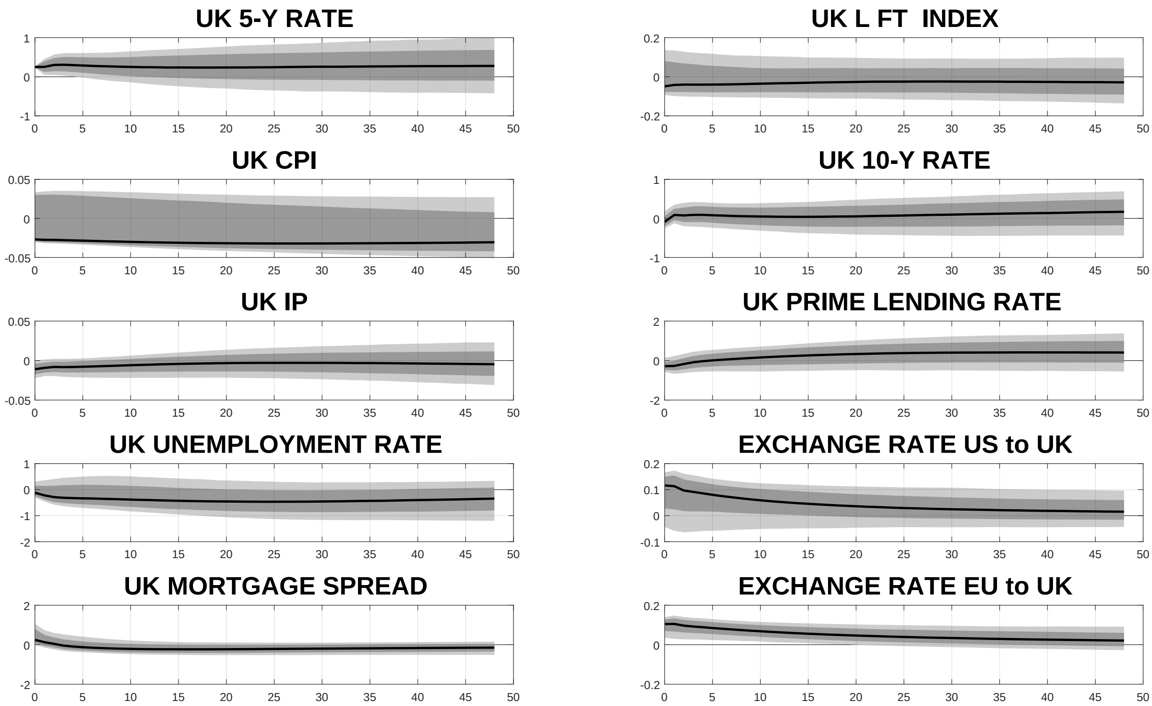
Note: US monetary policy surprise series of Gertler and Karadi (2015). Responses to a 25 basis point increase in the US treasury 1-year rate. A sign restriction is set on the response of the mortgage spread on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

policy shock of the ECB. Some recent papers provide evidence for heterogeneous effects across euro area countries of unconventional monetary policy shocks of the ECB on, for example, government bond yields, industrial production and CPI (see e.g., Ciccarelli et al., 2013; Georgiadis, 2015; Hachula et al., 2016; Boeckx et al., 2017; Burriel and Galesi, 2018).

3.6.2 International Monetary Policy Transmission

Figures 3.7, 3.8, and 3.9 show the impulse responses of exchange rates and domestic variables to monetary policy shocks for the US, UK, and EA in models (4), (5), and (6). Figure 3.10 gives the reactions of US, UK, and EA variables to a US monetary policy shock in model (7). Figures 3.11 and 3.12 show the impulse responses of foreign

Figure 3.8: Model (5): Responses of exchange rates and UK variables to a contractionary UK monetary policy shock



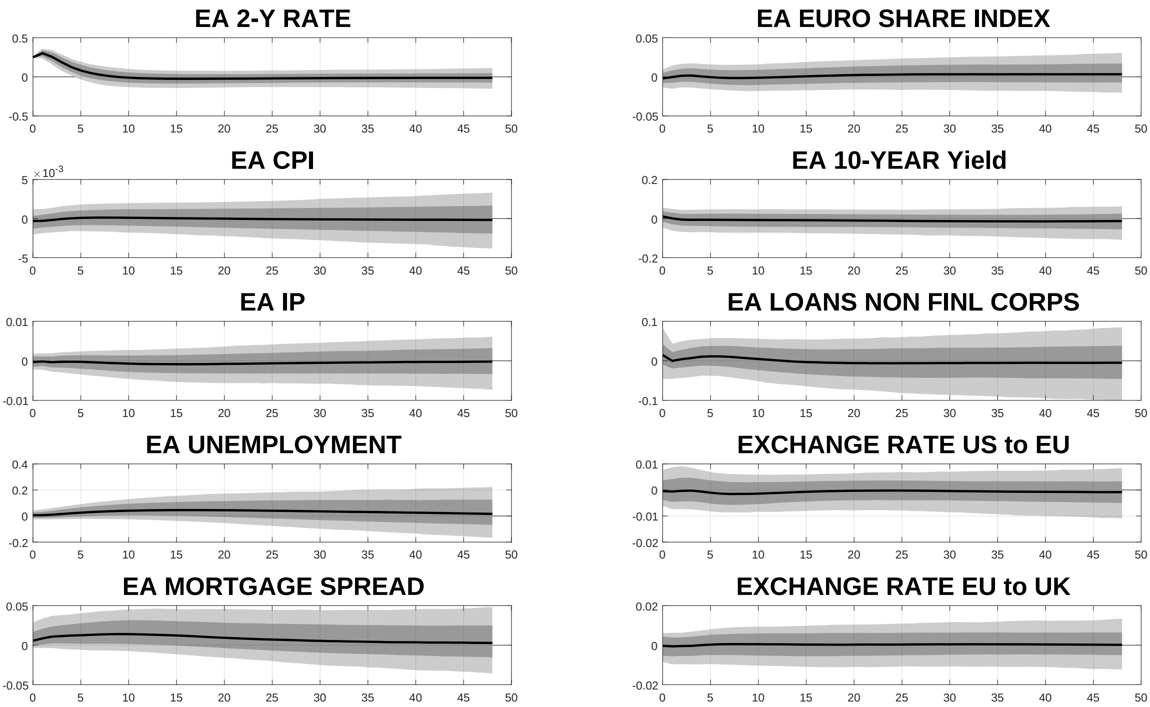
Note: UK monetary policy surprise series of Gerko and Rey (2017). Responses to a 25 basis point increase in the 5-year yield from British government securities. A sign restriction is set on the response of the mortgage spread on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1982 to March 2017. Proxy data from July 1997 to Jan 2015.

variables to UK and EA monetary policy shocks in models (8) and (9). The models are based on tight priors for the standard deviation of the measurement error in the proxy equation.

Our results provide some evidence for international spillovers via the exchange rate channel. The cross-border transmission of monetary policy shocks is asymmetric across the US, UK, and EA. A US monetary policy shock causes an appreciation of the US dollar with regard to the British pound and Euro. The British pound appreciates with regard to the Euro but not to the US dollar in reaction to a UK monetary policy shock. A EA monetary policy shock leads to no significant reaction of the exchange rates. We find no evidence of cross-border macroeconomic effects.

A monetary tightening in the US is transmitted to the EA and UK by the exchange rate channel. Exchange rates drop modestly on impact as shown in figure 3.7 (model 4).

Figure 3.9: Model (6): Responses of exchange rates and EA variables to a contractionary EA monetary policy shock



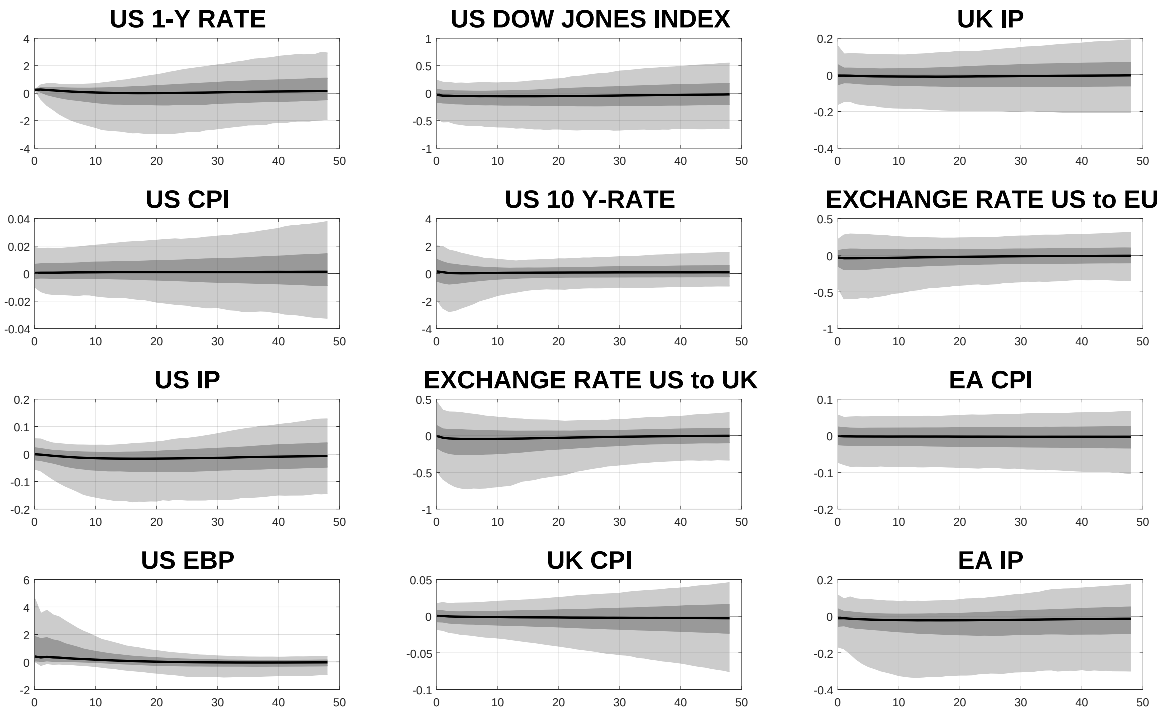
Note: EA monetary policy surprise series of Rogers et al. (2017). Responses to a 25 basis point increase in the Italian gross yield of benchmark 5-year rate. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 2000 to Oct 2017. Proxy data from Oct 2007 to Dec 2015.

The dollar appreciates with regard to the British pound and the Euro. The response gets insignificant for both exchange rates after ten months. Rogers et al. (2017) show that as a reaction to an expansionary US monetary policy shock the dollar depreciates with respect to foreign currencies. Gerko and Rey (2017) find an appreciation effect of the US dollar to the British pound as response to a monetary tightening.

A UK monetary policy shock spills-over to the EA via exchange rates. A 25 basis point increase in the UK 5-year yield increases the exchange rate of the Euro to the British pound as shown in figure 3.8 (model 5). The British pound appreciates with regard to the Euro. The exchange rate of the US dollar to the British pound raises on impact. The response is only statistically significant at the 68% confidence bands. Gerko and Rey (2017) also find no significant appreciation effect of the British pound with regard to the US dollar for a UK monetary policy shock.

A monetary policy shock of the ECB causes no significant exchange rate reaction as

Figure 3.10: Model (7): Responses of US, UK, and EA variables to a contractionary US monetary policy shock



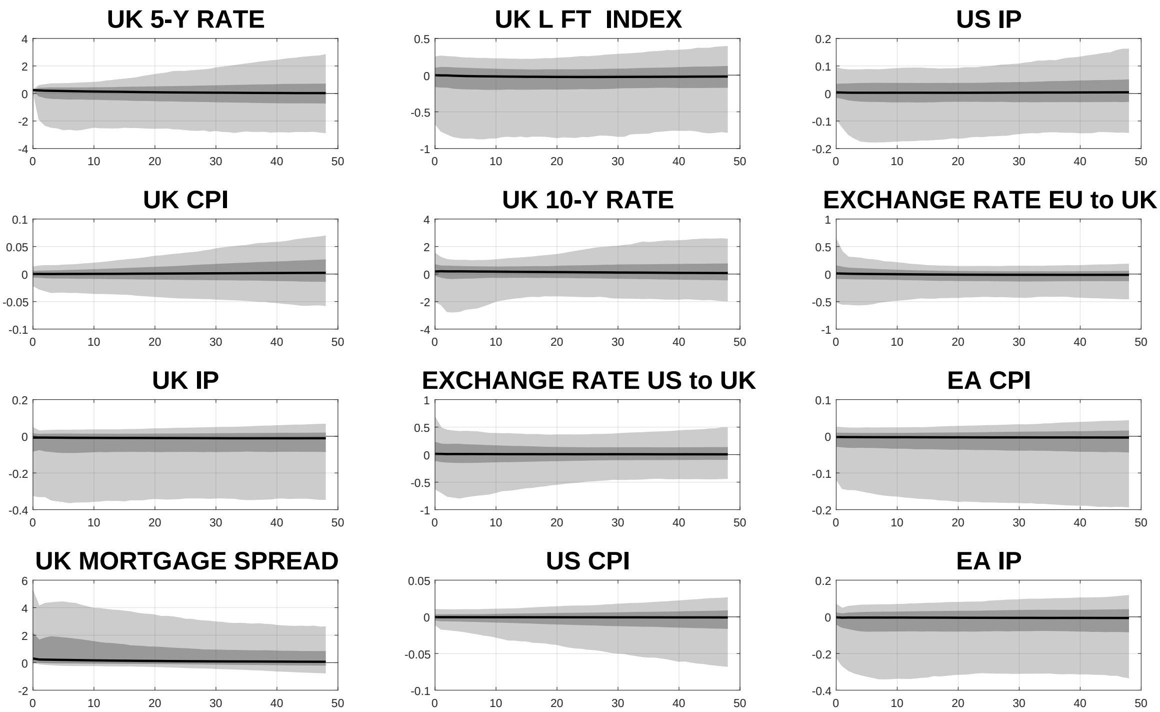
US monetary policy surprise series of Gertler and Karadi (2015). Responses of US, UK, and EA variables to a 25 basis point increase in the US treasury 1-year rate. A sign restriction is set on the response of the mortgage spread on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1991 to June 2017. Proxy data from Jan 1991 to June 2012.

shown in figure 3.9 (model 6). As explained above, this finding can be driven by the identification strategy or be explained by the aggregated EA level. However, adding the exchange rates to the single country VAR models leads to mainly insignificant reactions of the domestic variables in all three models.

We find no evidence of macroeconomic spillovers across the US, UK, and EA.[18] Figure 3.10 shows the responses of US, UK, and EA variables to a contractionary US monetary policy shock (model 7). Foreign consumer prices and industrial production do not respond significantly. In contrast to the previous findings, the US monetary policy shock does not cause a significant reaction of US variables and exchange rates. Confidence bands are considerably larger than in the single country VAR models. Fig-

[18]Further results for three-country VAR models are given in 3.F.5.

Figure 3.11: Model (8): Responses of US, UK, and EA variables to a contractionary UK monetary policy shock



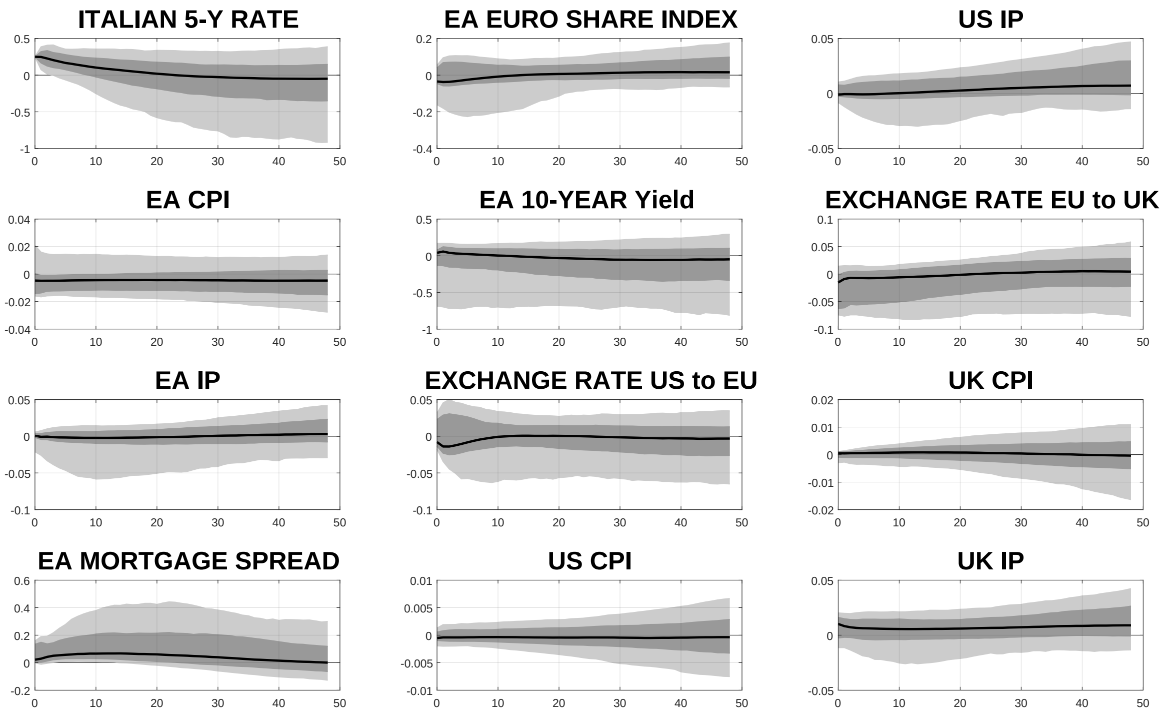
Note: UK monetary policy surprise series of Gerko and Rey (2017). Responses of US, UK, and EA variables to a 25 basis point increase in the 5-year yield from British government securities. A sign restriction is set on the response of the mortgage spread on the monetary policy shock. The response it set to be positive. The gray shaded ares is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1991 to March 2017. Proxy data from July 1997 to Jan 2015.

ure 3.11 gives the responses of US, UK, and EA variables to a UK monetary policy shock (model 8). Figure 3.12 shows the reactions of US, UK, and EA variables to a EA monetary policy shock (model 9). Again, no evidence for international spillovers is found. Adding more variables does not change the insignificance of the results. The insignificant reactions are further discussed in the next section 3.6.3.

3.6.3 Discussion of the Results

Regarding our results three issues have to be taken into account. First, to identify the monetary policy shocks, high relevance priors are necessary. Otherwise, the proxies seem to carry insufficient information for identification. Second, sample sizes for the proxy data and the data of the VAR model differ across VAR models and also across

Figure 3.12: Model (9): Responses of US, UK, and EA variables to a contractionary EA monetary policy shock



Note: EA monetary policy surprise series of Rogers et al. (2017). Responses of US, UK, and EA variables to a 25 basis point increase in the Italian gross yield of benchmark 5-year rate. The gray shaded ares is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 2000 to Oct 2017. Proxy data from Oct 2007 to Dec 2015.

single country models and the three-country models. Third, the chosen selection prior for the three-country model seems to only partly help to counter the high estimation uncertainty due to the relatively large number of variables included. These aspects might be the main reasons for finding insignificant results.

High relevance priors are needed to identify monetary policy shocks. The prior on the standard deviation of the measurement error of the proxy equation is crucial as it determines the relevance of the prior. This prior states a tight relation and thus that the prior is very informative about the structural shock. The prior set is $\sigma_{v,i} = 0.2 \times \sigma_{m_{i,t_{1m}:T_m}}$ for the US and UK proxy and $\sigma_{v,EA} = 0.4 \times \sigma_{m_{EA,t_{1m}:T_m}}$ for the EA proxy. In addition, while the prior for b_i is set to $\mathcal{N}(0, 0.01)$ for the US and UK, the prior is set to $\mathcal{N}(0.2, 0.0001)$ for the EA.

The relevance of the proxies is measured by $b_i^2/(b_i^2 + \sigma_{v,i}^2)$ (see also Caldara and

Table 3.3: Relevance measure based on posteriors for b_i and $\sigma_{v,i}$ with different priors

single country VAR		percentiles of relevance measure				
	prior on $\sigma_{v,i}$	0.05	0.16	0.50	0.86	0.95
(1) US	high relevance	0.19	0.64	0.77	0.81	0.84
(1) US	high relevance with sign rest	0.28	0.42	0.57	0.70	0.75
(1) US	inverse Gamma	0.00	0.00	0.02	0.05	0.08
(2) UK	high relevance	0.30	0.38	0.46	0.53	0.57
(2) UK	high relevance with sign rest	0.47	0.54	0.63	0.69	0.72
(2) UK	inverse Gamma	0.00	0.00	0.02	0.06	0.08
(3) EA	high relevance	0.97	0.97	0.97	0.97	0.97
(3) EA	inverse Gamma	0.02	0.03	0.04	0.06	0.07

Note: The relevance of the proxies is measured by $b_i^2/(b_i^2 + \sigma_{v,i}^2)$. The measure is given for different percentiles of the posteriors of b_i and $\sigma_{v,i}$. The high relevance prior is $\sigma_{v,i} = 0.2 \times \sigma_{m_{i,t_{1m}:T_m}}$ for the US and UK proxy and $\sigma_{v,EA} = 0.4 \times \sigma_{m_{EA,t_{1m}:T_m}}$ for the EA proxy.

Herbst, 2016). That is, given equation (3.5) the relevance measure is the squared correlation between $m_{i,t}$ and $e_{MPi,t}$. This value is comparable to the reliability indicator of the proxy of Mertens and Ravn (2013). The larger the relevance measure, the more informative is the proxy for the identification. Table 3.3 shows for the single country VAR models (1), (2), and (3) the relevance measure for the high relevance prior and the inverse Gamma prior on $\sigma_{v,i}$ calculated with the posterior distributions for b_i and $\sigma_{v,i}$. In mean, the relevance measure is 0.77 for high relevance prior for the US, 0.46 for the UK, and 0.97 for the EA. These values are considerably higher than the relevance values for the inverse Gamma prior. Here, the measure ranges in mean between 0.02 and 0.04. Thus, without the high relevance priors, the proxies carry very low information for the identification of the monetary policy shocks. Including the sign restriction increases the relevance for the UK. 3.E gives the posterior distributions of b_i for the high relevance prior and the inverse Gamma prior on $\sigma_{v,i}$.

However, the high relevance prior for the US and UK seem to cause puzzling reaction contradicting economic theory as shown in 3.F.1. That is why, an additional sign restriction is set on the US excess bond premium or the UK mortgage spread to absorb the effects of the high relevance prior.

Furthermore, the models are based on different time samples. In particular, the proxy data for the EA cover a considerably shorter period than the US and UK proxies. The data range for the EA variables is also shorter than for the other two countries. This has to be taken into account when comparing results across countries.

Moreover, by including several variables in the three-country VAR models the estimation uncertainty increases. The selection prior on the reduced form coefficients is chosen to counter the higher uncertainty by reducing the dimension of the model through setting zero restrictions. However, the insignificant results for international monetary policy transmission could be an indication that the high estimation uncertainty cannot be offset by the selection prior in the form employed in this paper. It is important to note that this fact is not changed by the information on reduced-form coefficients coming from our proxies. Even if we shut down this channel by accepting every draw from the normal-Wishart posterior distribution, the impulse-responses in the multi-country model remain insignificant. An alternative explanation could be that the multi-country VAR model allows to correctly model a high number of channels for the transmission of monetary policy. If these channels counteract and partly cancel each other, the overall effect of monetary policy shocks may indeed be insignificant. However, this explanation is unlikely, as at least some of the channels should be visible in a significant reaction of financial market variables even if aggregate macroeconomic effects are insignificant.

3.7 Conclusions

In the context of highly interlinked financial and real markets the effects of cross-border transmission of monetary policy are important to analyze. Unintended harmful spillover effects and increased global instabilities may call for the need of policy coordination. The start of the zero lower bound period and the implementation of unconventional monetary policy measures of central banks rise the question of the international effects of unconventional monetary policy. While recent literature provides evidence for financial spillover, the macroeconomic effects are less clear.

We analyze empirically the international transmission of monetary policy shocks of the US, UK, and EA. We use a Bayesian proxy three-country structural vector autoregressive model. The model captures interdependencies across the economies as it jointly models variables of all countries in one model. The three-country VAR model allows us to trace the dynamic cross-border effects of monetary policy shocks. We use external instruments constructed from high-frequency government bond futures to identify the monetary policy shocks. With this identification strategy we avoid the recursive identification scheme commonly employed in multi-country models, which is questionable as there is no clear theoretical indication regarding the ordering of variables in panel VAR models.

Our findings provide no evidence of strong cross-border transmission of monetary policy shocks. First, international spillovers from monetary policy shocks seem to have no substantial effect on macroeconomic variables globally. Second, there are some asymmetries between the effects of the spillover of US monetary policy shocks and surprises of the UK and EA. While a US monetary policy shock causes a modest reaction of exchange rates to the British pound and Euro, the UK monetary policy shock affects only the exchange rate to the Euro. A EA monetary policy shock does not spill-over to exchange rates. Third, domestic monetary policy shocks lead to drops in domestic stock prices for all three regions.

The insignificance of our results in the multi-country VAR model point to deficiencies in the estimation and shock identification of panel VAR models. It is not yet clear which endogenous variables are necessary to correctly specify a multi-country VAR model and how prior distributions (specifically, the selection prior and the prior distribution for $\sigma_{v,i}$) need to be chosen to counter the estimation uncertainty arising from parameter proliferation. This is an area for future research.

Appendix

3.A Posterior Distributions

The posterior distributions of b_i and $\sigma_{v,i}$ are obtained by:

$$\begin{aligned} p(b_i, \sigma_{v,i}|y_{1:T_y}, m_{i,t_{1m}:T_m}, \Phi, \Sigma, \Omega) =& p(b_i)p(\sigma_{v,i})p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi, \Sigma, \Omega, b_i, \sigma_{v,i}) \\ \propto& V_{b,i}^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(b_i - b_{0,i})V_{b,i}^{-1}(b_i - b_{0,i})\right\} \\ &\times \sigma_{v,i}^{-\frac{s_i}{2}-1} exp\left\{-\frac{S_{v,i}(s_i-2)}{2\sigma_{v,i}}\right\} \\ &\times \sigma_{v,i}^{-\frac{T}{2}} exp\left\{-\frac{1}{2}(m_{i,t} - \mu_{m,i})\sigma_{v,i}^{-1}(m_{i,t} - \mu_{m,i})'\right\} \\ \propto& V_{b,i}^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(b_i - \bar{b}_i)\bar{V}_{b,i}^{-1}(b_i - \bar{b}_i)\right\} \\ &\times \sigma_{v,i}^{-\frac{\bar{s}_i}{2}-1} exp\left\{-\frac{\bar{S}_{v,i}(\bar{s}_i-2)}{2\sigma_{v,i}}\right\}, \end{aligned}$$

where

$$\begin{aligned} \bar{b}_i &= \bar{V}_{b,i}(V_{b,i}^{-1}b_{0,i} + \sigma_{v,i}^{-1}m_{i,t}e'_{i,t}) \\ \bar{V}_{b,i} &= (V_{b,i}^{-1} + \sigma_{v,i}^{-1}e_{i,t}e'_{i,t})^{-1} \\ \bar{s}_i &= s_i + T \\ \bar{S}_{v,i} &= S_{v,i}(s_i-2) + (m_{i,t} - b_i e_{i,t})(m_{i,t} - b_i e_{i,t})'. \end{aligned}$$

Thus, the posterior distributions of b_i is a normal distribution and for $\sigma_{v,i}$ an inverse Gamma distribution given by $b_i \sim \mathcal{N}(\bar{b}_i, \bar{V}_{b,i})$ and $\sigma_{v,i} \sim \mathcal{IG}(\bar{s}_i, \bar{S}_{v,i})$. Note that the distributions of b_i and $\sigma_{v,i}$ still depend on the reduced form parameters and Ω due to e_i.

3.B Estimation Algorithm

We estimate the Bayesian proxy three-country SVAR using a Metropolis-within-Gibbs algorithm. The algorithm has the following steps at iteration *iter*:

1. Draw reduced form parameters, $\Phi, \Sigma|y_{1:T_y}, m_{i,t_{1m}:T_m}, \Omega^{old}, b_i^{old}, \sigma_{v,i}^{old}$.
    a) Draw Σ^{new}
    b) Draw Φ^{new}

c) If $\alpha \leq a$, accept new draw and set $\Phi^{iter} = \Phi^{new}$ and $\Sigma^{iter} = \Sigma^{new}$. Otherwise set $\Phi^{iter} = \Phi^{old}$ and $\Sigma^{iter} = \Sigma^{old}$. The probability α is given by

$$\alpha = min\left\{\frac{p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi^{new}, \Sigma^{new}, \Omega^{new}, b_i^{old}, \sigma_{v,i}^{old})p(\Sigma^{new})q(\Sigma^{old}|\Sigma^{new})}{p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi^{old}, \Sigma^{old}, \Omega^{old}, b_i^{old}, \sigma_{v,i}^{old})p(\Sigma^{old})q(\Sigma^{new}|\Sigma^{old})}, 1\right\}$$

and $a \sim U(0,1)$.

2. Draw the rotation matrix $\Omega|y_{1:T_y}, m_{i,t_{1m}:T_m}, \Omega^{old}, b_i^{old}, \Sigma^{old}$

a) Draw an independent standard normal matrix of dimension $[G \times G]$ and take the QR-decomposition of this matrix to get Ω.

b) If $\alpha \leq a$, accept new draw and set $\Omega^{iter} = \Omega^{new}$. Otherwise set $\Omega^{iter} = \Omega^{old}$. The probability α is given by

$$\alpha = min\left\{\frac{p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi^{iter}, \Sigma^{iter}, \Omega^{new}, b_i^{old}, \sigma_{v,i}^{old})p(\Sigma^{new})q(\Sigma^{old}|\Sigma^{new})}{p(m_{i,t_{1m}:T_m}|y_{1:T_y}, \Phi^{iter}, \Sigma^{iter}, \Omega^{old}, b_i^{old}, \sigma_{v,i}^{old})p(\Sigma^{old})q(\Sigma^{new}|\Sigma^{old})}, 1\right\}$$

and $a \sim U(0,1)$.

3. Draw b_i of $b_i|y_{1:T_y}, m_{i,t_{1m}:T_m}, \Omega^{iter}, b_i^{old}, \sigma_{v,i}^{old}$ from $\mathcal{N}(\bar{b}_i, \bar{V}_{b,i})$.

4. Draw $\sigma_{v,i}|y_{1:T_y}, m_{i,t_{1m}:T_m}, \Omega^{iter}, b_i^{old}, \sigma_{v,i}^{old}$ from $\mathcal{IG}(\bar{s}_i, \bar{S}_{v,i})$.

Step (2) updates Ω based on the ratio of the conditional likelihoods. We accept a new draw of Ω with a higher probability if it leads to a closer scaled version of the structural shock.

The algorithm allows the estimation based on different lengths of time series of $y_{1:T_y}$ and $m_{i,t_{1m}:T_m}$. The matrices Φ and Σ are drawn based on longer time series, while the update step in (1c) and the draws of Ω, b_i, and $\sigma_{v,i}$ can use a shorter sample determined by the availability of the proxy data.

Furthermore, the algorithm easily allows for the inclusion of additional restrictions such as sign or magnitude restrictions. These additional restrictions can be checked for in step (2).

3.C Data

Table 3.4: Variables of the United States

Variables	**Description and Source**
US CPI	US consumer price index, all items less food and energy (seasonally adjusted (SA), logarithm), Source: US Bureau of Labor Statistics
US IP	US industrial production index (SA, logarithm), Source: Federal Reserve
US UNEMPLOYMENT RATE	US unemployment rate (SA, logarithm), Source: US Bureau of Labor Statistics
US 1-Y RATE	US Treasury yield adjusted to constant maturity, 1 year, Source: Federal Reserve
US 10-Y RATE	US Treasury yield adjusted to constant maturity, 10 year, Source: Federal Reserve
US PRIME RATE BANKS	US prime rate charged by banks (monthly average), Source: Federal Reserve
US EBP	excess bond premium, Source: Federal Reserve
US DOW JONES INDEX	Dow Jones industrial share price index (SA, logarithm), Source: Reuters
EXCHANGE RATE US to EU	exchange rate US dollar to 1 Euro, Source: Datastream
EXCHANGE RATE US to UK	exchange rate US dollar to 1 British pound, Source: Datastream

Table 3.5: Variables of the United Kingdom

Variables	**Description and Source**
UK CPI	UK consumer price index all items (SA, logarithm), Source: Main Economic Indicators, OECD
UK IP	UK industrial production index (SA, logarithm), Source: ONS
UK UNEMPLOYMENT RATE	UK unemployment rate (SA, logarithm), Source: International Financial Statistics (IMF)
UK 5-Y RATE	yield from British government securities, 5 year nominal zero coupon, Source: Bank of England
UK 10-Y RATE	British government securities, 10 year nominal zero coupon, Source: Bank of England
UK PRIME LENDING RATE	UK major banks prime lending rate, Source: Reuters
UK MORTGAGE SPREAD	Variable rate mortgage spread over Bank Rate, Source: A millennium of macroeconomic data for the UK, BoE
UK L FT INDEX	UK FT all share index (SA, logarithm), Source: Reuters

Table 3.6: Variables of the euro area

Variables	**Description and Source**
EA CPI	EA consumer price index, all items (SA, logarithm), Source: European Central Bank
EA IP	EA industrial production index (SA, logarithm), Source: Eurostat
EA UNEMPLOYMENT RATE	EA unemployment rate (SA, logarithm), Source: European Central Bank
ITALIEN 5-Y RATE	Italian gross yield of benchmark 5-year BTP, Source: Bank of Italy
EA 2-Y RATE	EA government bond yield, 2 year, Source: European Central Bank
EA 10-YEAR Yield	EA government bond yield, 10 year, Source: European Central Bank
EA LOANS NON FINL CORPS	EA loans to non financial corporations over one million Euro, 1 to 5 years, Source: European Central Bank
EA MORTGAGE SPREAD	EA mortgage rate minus 3-month Euribor, Source: European Central Bank
EA EURO SHARE INDEX	EM Datastream Euro share price index (SA, logarithm), Source: Datastream
EXCHANGE RATE EA to UK	exchange rate Euro to 1 British pound, Source: Datastream

3.D Prior Choices

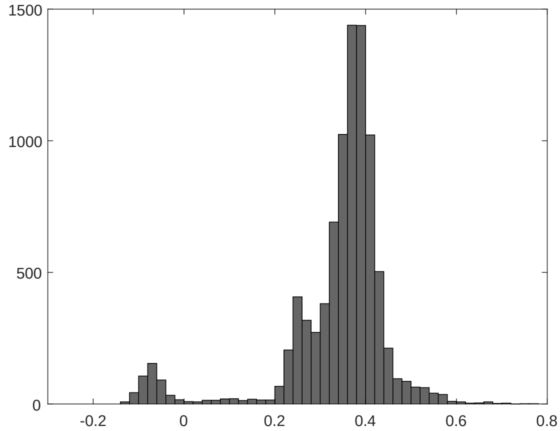
The selection hyperparameter γ in the selection prior for the reduced from coefficients is drawn from a Bernoulli distribution with prior probability of 0.5. Thus, under the prior setting a zero restriction or no zero restriction is equally likely. The parameter ρ of the mixture distribution for Σ is set to 0.9, d is set equal to the number of observations. The overall tightness of the Minnesota prior is set to 0.2, the parameter scaling down the variance for the coefficients of a distant lag to 3.5, the relative tightness of the constant is set to 0.5, and the hyperparameter for the covariance between coefficients is set to 0.5.

The prior for b_i is set to $\mathcal{N}(0, 0.01)$ for the US and UK and $\mathcal{N}(0.2, 0.0001)$ for the EA. The high relevance prior for $\sigma_{v,i}$ is $\sigma_{v,i} = 0.2 \times \sigma_{m_{i,t_{1m}:T_m}}$ for the US and UK and

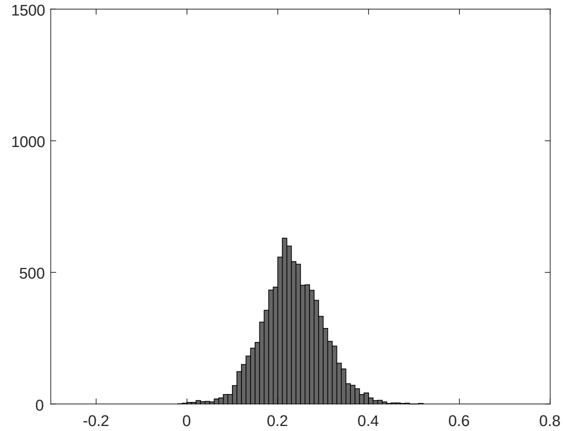
$\sigma_{v,i} = 0.4 \times \sigma_{m_{i,t_{1m}:T_m}}$ for the EA. The less informative prior for $\sigma_{v,i}$ is an inverse gamma distribution with degrees of freedom 2 and mean 0.02.

3.E Relevance of the Proxies

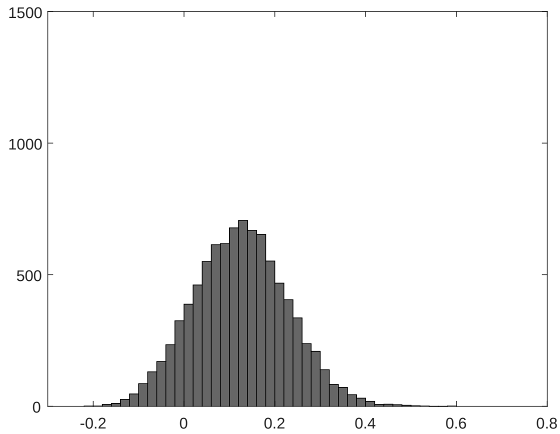
Figure 3.13: Model (1): Posterior distributions of b_{US} with different priors for $\sigma_{v,US}$



(a) high relevance prior



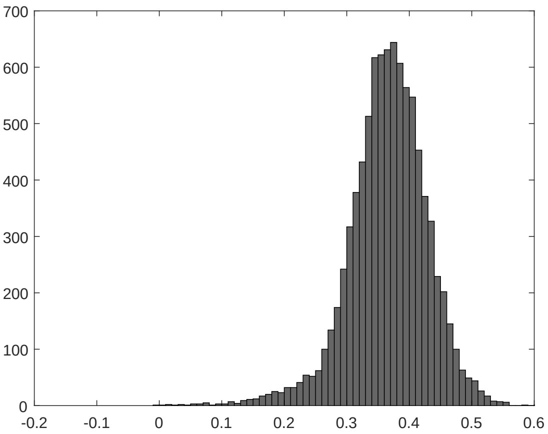
(b) high relevance prior and sign restriction



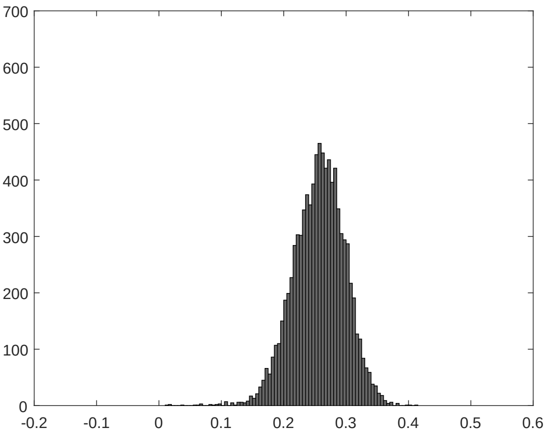
(c) inverse Gamma prior

Note: The posterior distributions for b_{US} with a tight prior for $\sigma_{v,US}$ and with a loose prior for $\sigma_{v,US}$. The model uses the US monetary policy surprise series of Gertler and Karadi (2015).

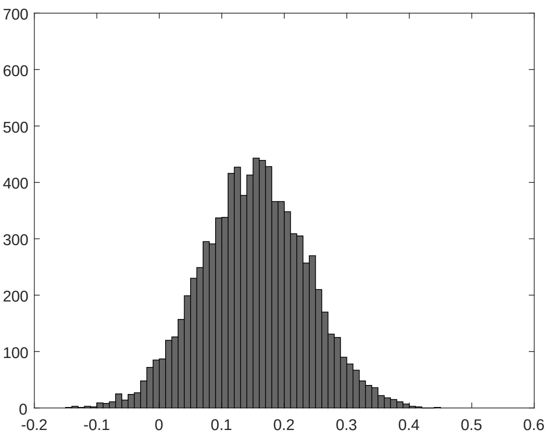
Figure 3.14: Model (2): Posterior distributions of b_{UK} with different priors for $\sigma_{v,UK}$



(a) high relevance prior



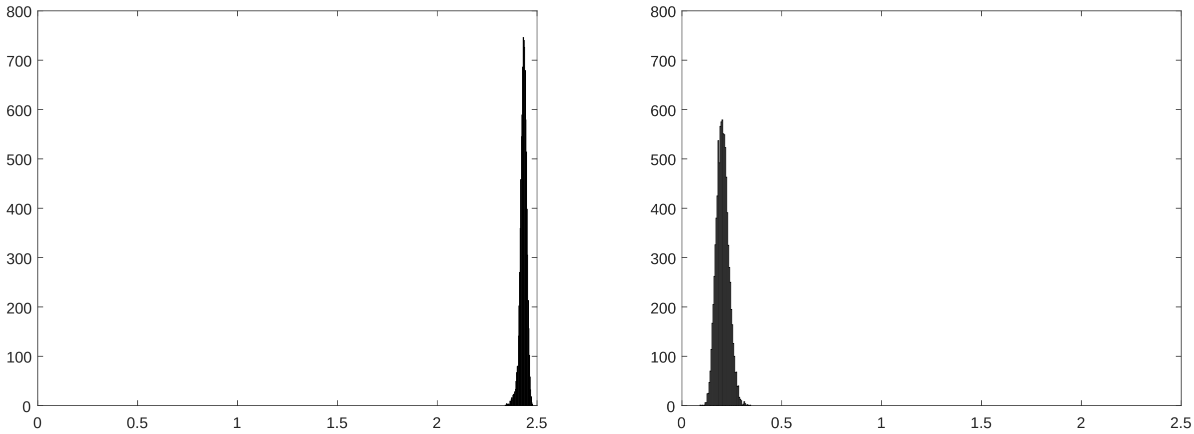
(b) high relevance prior and sign restriction



(c) inverse Gamma prior

Note: The posterior distributions for b_{UK} with a tight prior for $\sigma_{v,UK}$ and with a loose prior for $\sigma_{v,UK}$. The model uses the UK monetary policy surprise series of Gerko and Rey (2017).

Figure 3.15: Model (3): Posterior distributions of b_{EA} with different priors for $\sigma_{v,EA}$



(a) high relevance prior

(b) inverse Gamma prior

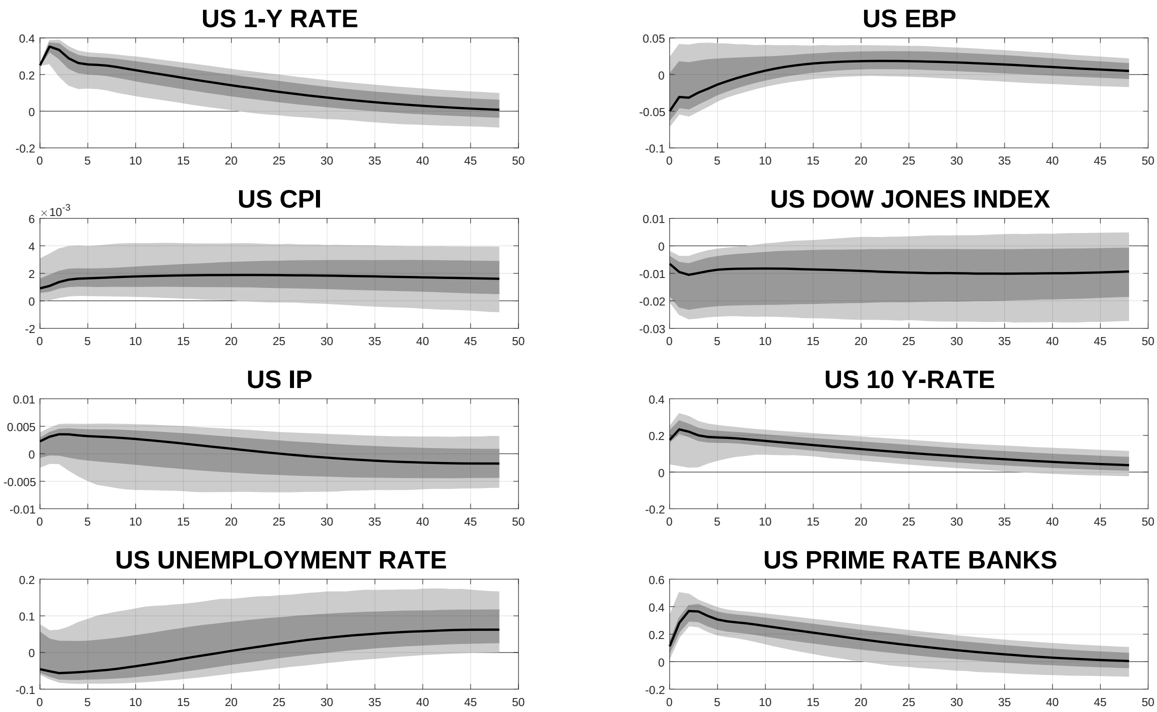
Note: The posterior distributions for b_{EA} with a tight prior for $\sigma_{v,EA}$ and with a loose prior for $\sigma_{v,EA}$. The model uses the EA monetary policy surprise series of Rogers et al. (2017).

3.F Additional Proxy VAR Results

3.F.1 Results with High Relevance Prior for US and UK without Additional Sign Restriction on Credit Spread Variable

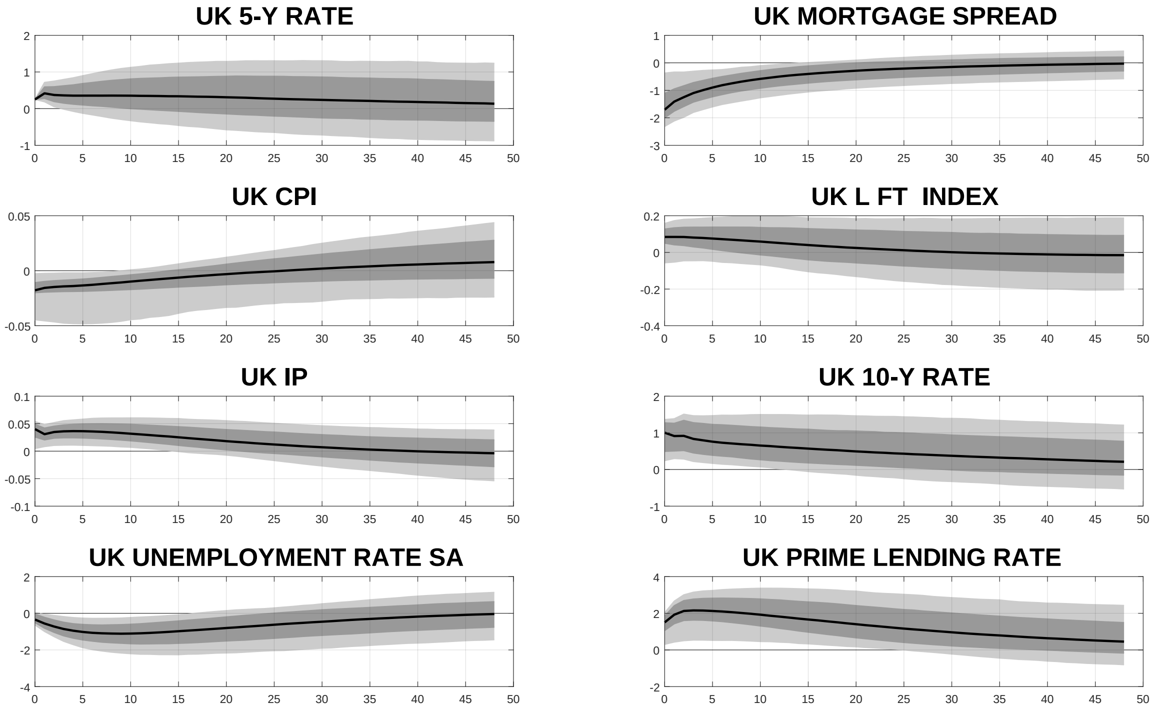
3.F.1 shows the impulse response functions for the single country VAR models for the US and UK with high relevance prior and without additional positive sign restrictions on US EBP and UK mortgage spread. No sign restriction on the excess bond premium leads to a price puzzle for the US. Figure 3.16 shows a significant positive reaction of US CPI while the excess bond premium drops on impact. This response is insignificant. The reaction of the other variables are comparable to the responses shown in figure 3.4. Without an additional sign restrictions some impulse responses of the UK model contradict theoretical arguments. IP increases on impact and the unemployment rate decreases, figure 3.17. The mortgage spread drops significantly. In contrast to figure 3.5 the stock price index raises on impact but the reaction is not significant.

Figure 3.16: Responses of US variables to a contractionary US monetary policy shock without sign restriction



Note: US monetary policy surprise series of Gertler and Karadi (2015). Responses of US variables to a 25 basis point increase in the US treasury 1-year rate. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

Figure 3.17: Responses of UK variables to a contractionary UK monetary policy shock without sign restriction

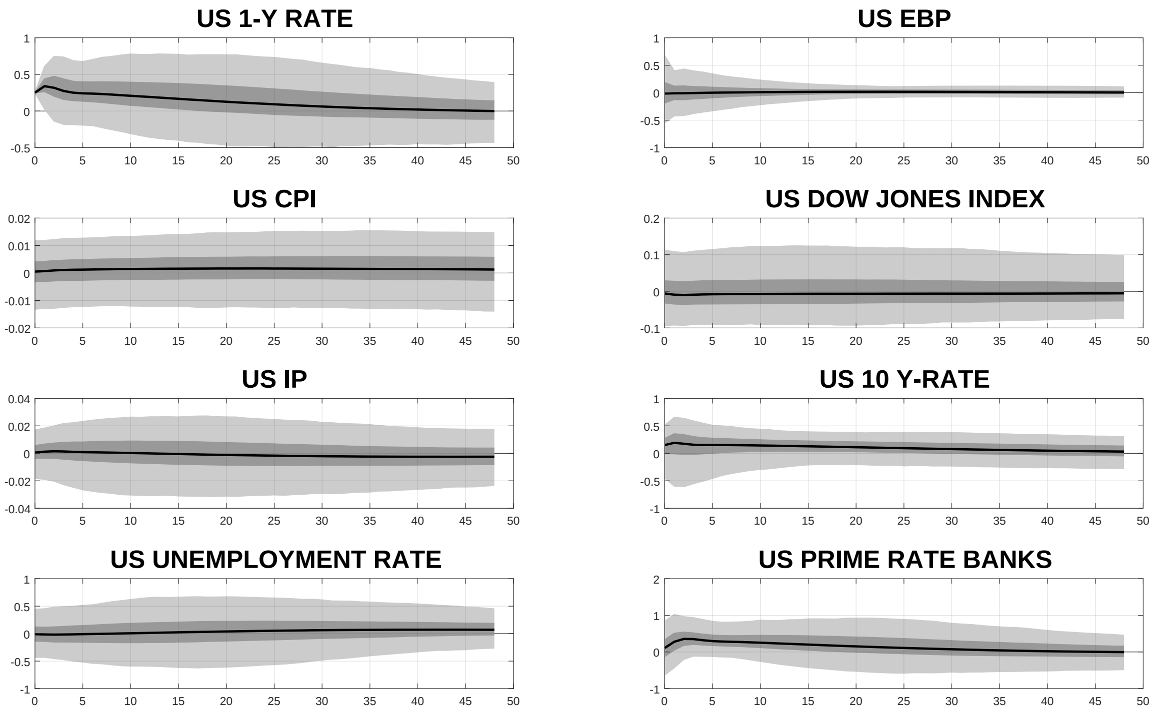


Note: UK monetary policy surprise series of Gerko and Rey (2017). Responses of UK variables to a 25 basis point increase in the 5-year yield from British government securities. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1982 to March 2017. Proxy data from July 1997 to Jan 2015.

3.F.2 Results with Inverse Gamma Prior

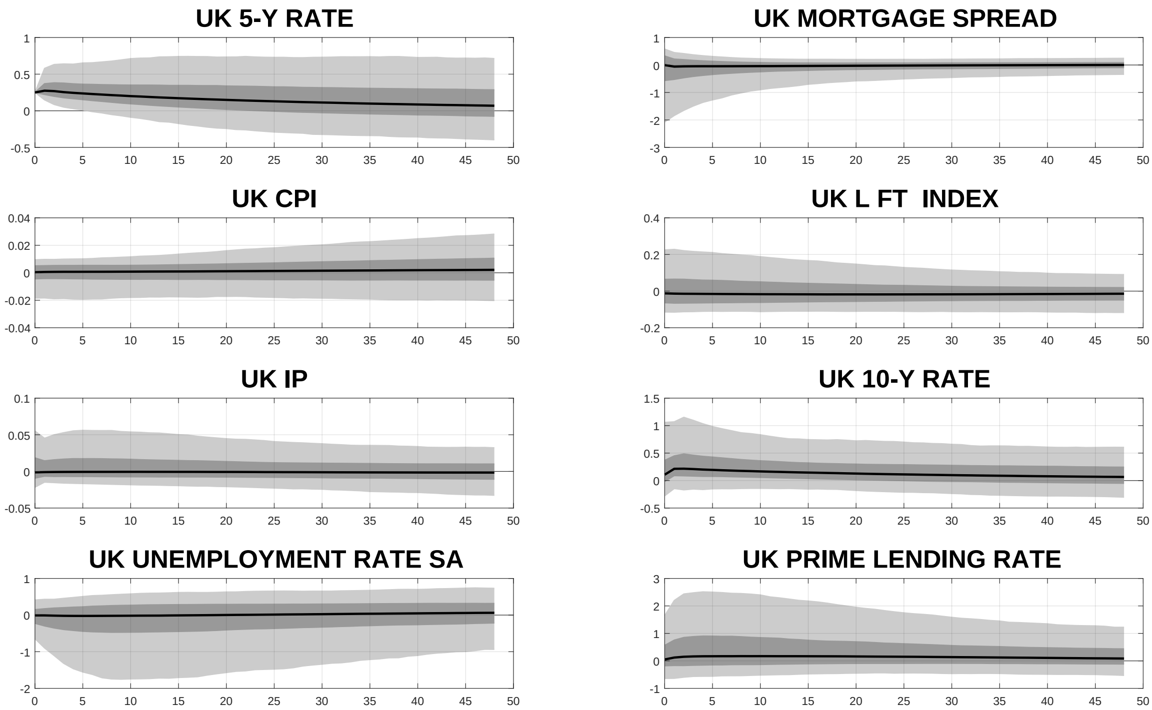
3.F.2 shows the impulse response functions for the single country VAR models for the US, UK and EA. The prior on the standard deviation of the measurement error, $\sigma_{v,i}$, in the proxy equation is assumed to be inverse Gamma with degrees of freedom of 2 and centering coefficient 0.02. The distribution for the standard deviation of the measurement error determines the tightness of the relationship of the proxy and the SVAR. This prior specification is not very informative.

Figure 3.18: Responses of US variables to a contractionary US monetary policy shock - inverse Gamma prior



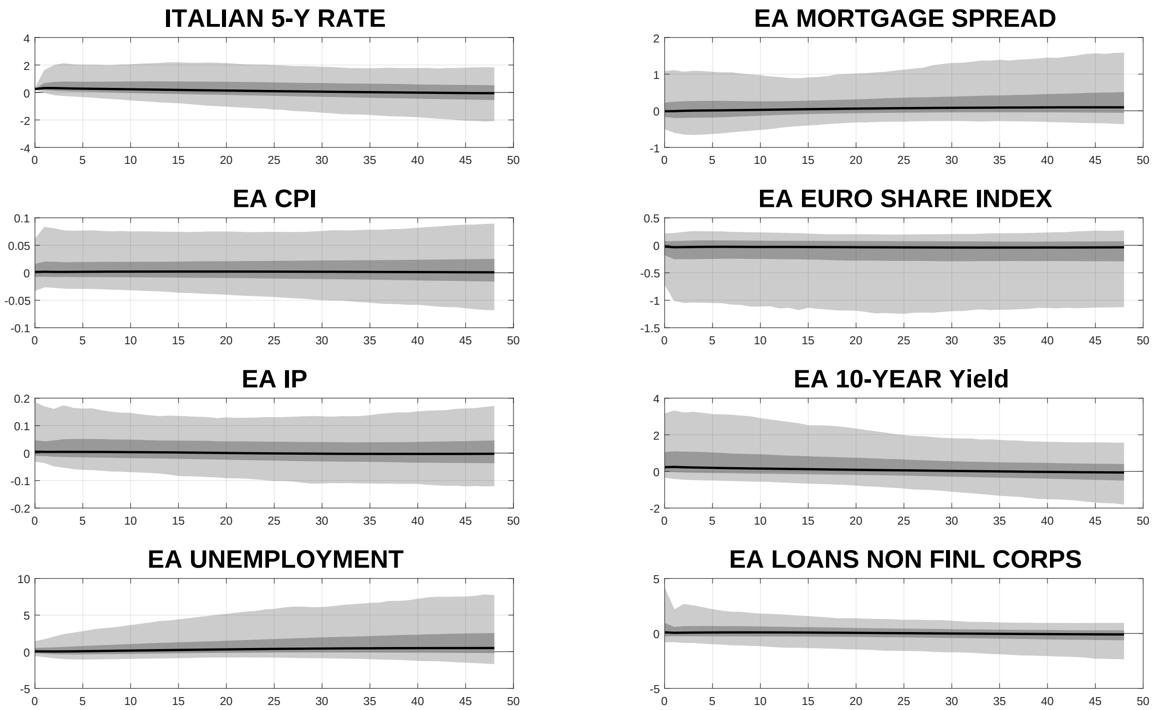
Note: US monetary policy surprise series of Gertler and Karadi (2015). For the proxy equation a loose prior is used for the standard deviation of the measurement error, $\sigma_{v,i} \sim \mathcal{IG}$. Responses of US variables to a 25 basis point increase in the US treasury 1-year rate. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

Figure 3.19: Responses of UK variables to a contractionary UK monetary policy shock - inverse Gamma prior



Note: UK monetary policy surprise series of Gerko and Rey (2017). For the proxy equation a loose prior is used for the standard deviation of the measurement error, $\sigma_{v,i} \sim \mathcal{IG}$. Responses of UK variables to a 25 basis point increase in the 5-year yield from British government securities. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1982 to March 2017. Proxy data from July 1997 to Jan 2015.

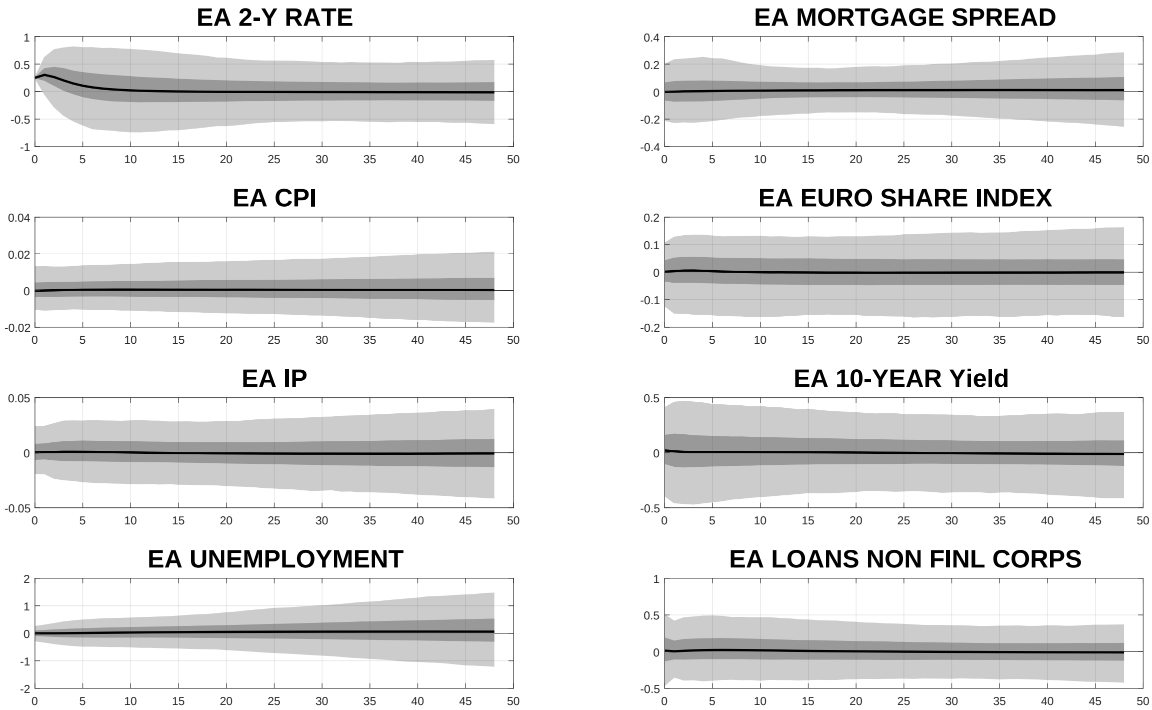
Figure 3.20: Responses of EA variables to a contractionary EA monetary policy shock - inverse Gamma prior



Note: EA monetary policy surprise series of Rogers et al. (2017). For the proxy equation a loose prior is used for the standard deviation of the measurement error, $\sigma_{v,i} \sim \mathcal{IG}$. Responses of EA variables to a 25 basis point increase in the Italian gross yield of benchmark 5-year rate. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 2000 to Oct 2017. Proxy data from Oct 2007 to Dec 2015.

3.F.3 Results for the EA with Different Monetary Policy Indicator

Figure 3.21: Responses of EA variables to a contractionary EA monetary policy shock - EA government bond 2-year rate as policy indicator

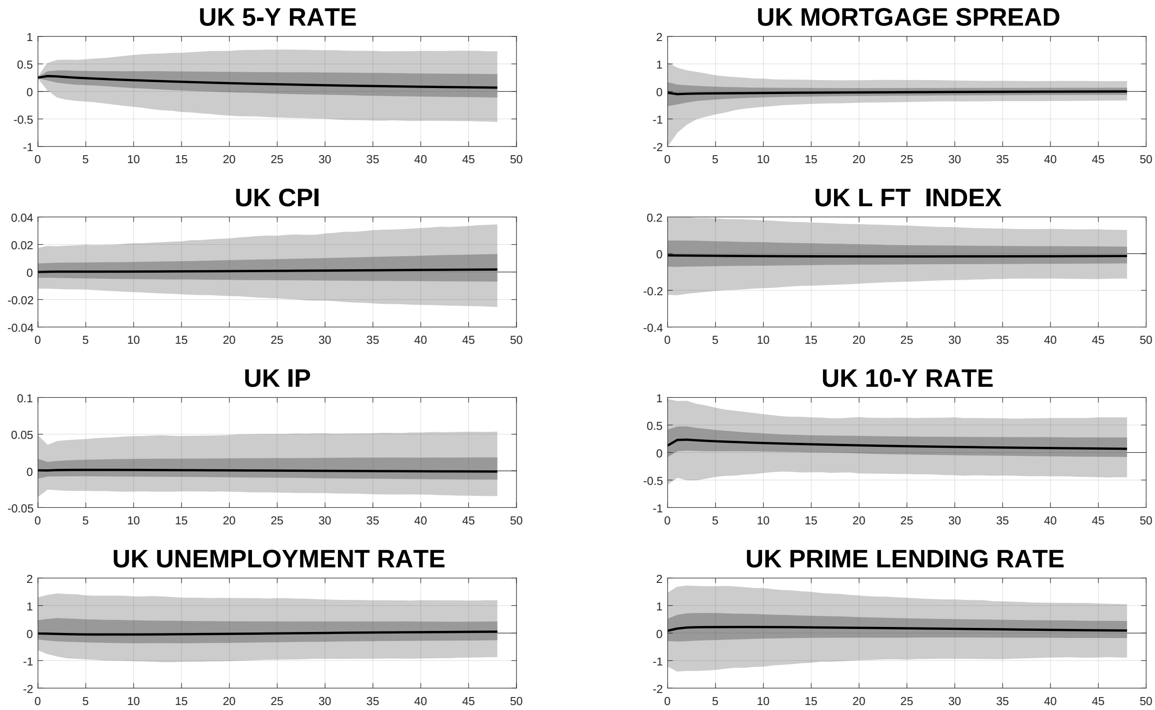


Note: EA monetary policy surprise series of Rogers et al. (2017). Responses of EA variables to a 25 basis point increase in the EA government bond 2-year rate. The prior on the standard deviation of the measurement error, $\sigma_{v,i}$, in the proxy equation is assumed to be inverse Gamma with degrees of freedom of 2 and centering coefficient 0.02. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 2000 to Oct 2017. Proxy data from Oct 2007 to Dec 2015.

3.F.4 Results for the UK with Different Proxies

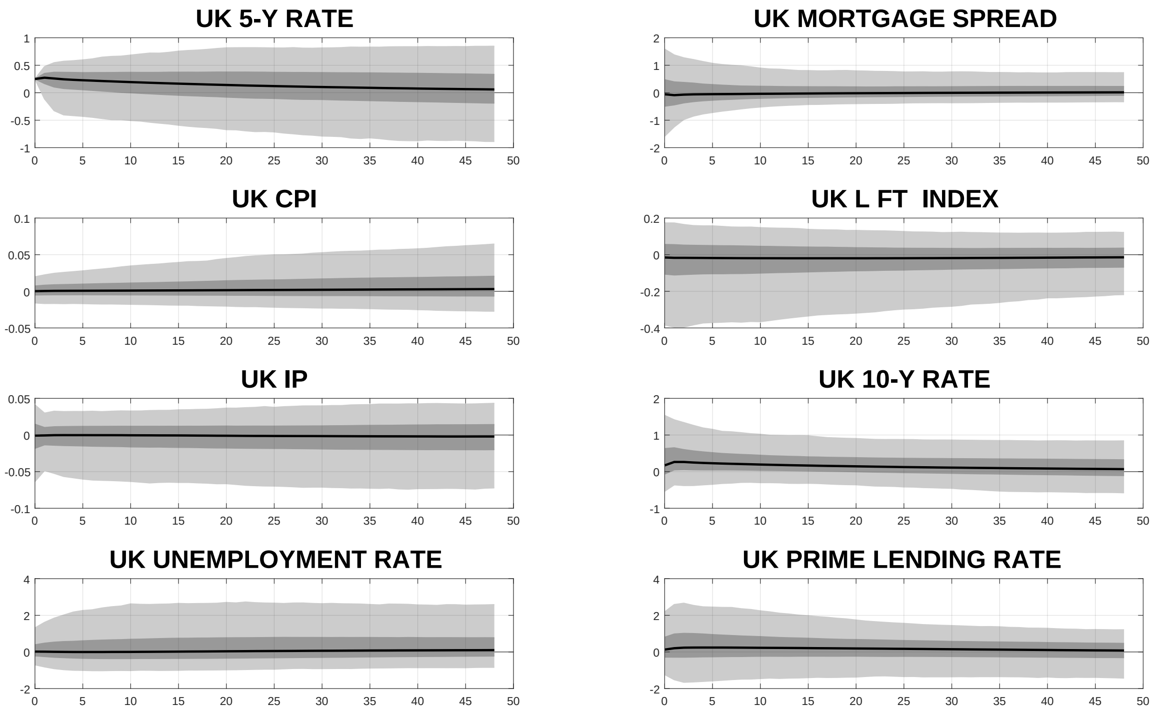
Figure 3.22 gives the responses using the monetary policy surprise series of Rogers et al. (2017). This proxy does not distinguish between rate and minutes and inflation report announcements. Figure 3.23 uses the monetary policy surprises around rate announcements of Gerko and Rey (2017) as proxy. Thus, it captures the changes in the target rate but does not include the announcements additionally made in minutes and inflation reports.

Figure 3.22: Responses of UK variables to a contractionary UK monetary policy shock - proxy of Rogers et al. (2017)



Note: UK monetary policy surprise series of Rogers et al. (2017). Responses of UK variables to a 25 basis point increase in the 5-year yield from British government securities. The prior on the standard deviation of the measurement error, $\sigma_{v,i}$, in the proxy equation is assumed to be inverse Gamma with degrees of freedom of 2 and centering coefficient 0.02. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1982 to March 2017. Proxy data from Oct 2008 to Dec 2015.

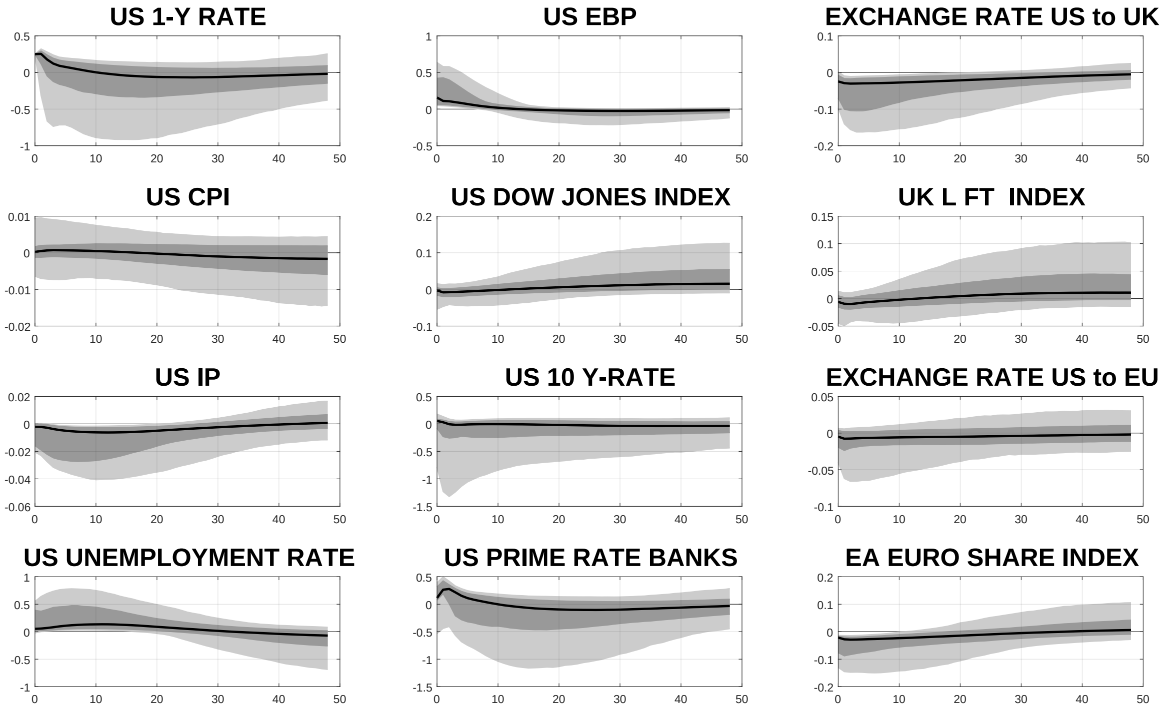
Figure 3.23: Responses of UK variables to a contractionary UK monetary policy shock - proxy around rate announcements of Gerko and Rey (2017)



Note: UK monetary policy surprise series around rate announcements of Gerko and Rey (2017). Responses of UK variables to a 25 basis point increase in the 5-year yield from British government securities. The prior on the standard deviation of the measurement error, $\sigma_{v,i}$, in the proxy equation is assumed to be inverse Gamma with degrees of freedom of 2 and centering coefficient 0.02. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR data from Jan 1982 to March 2017. Proxy data from July 1997 to Jan 2015.

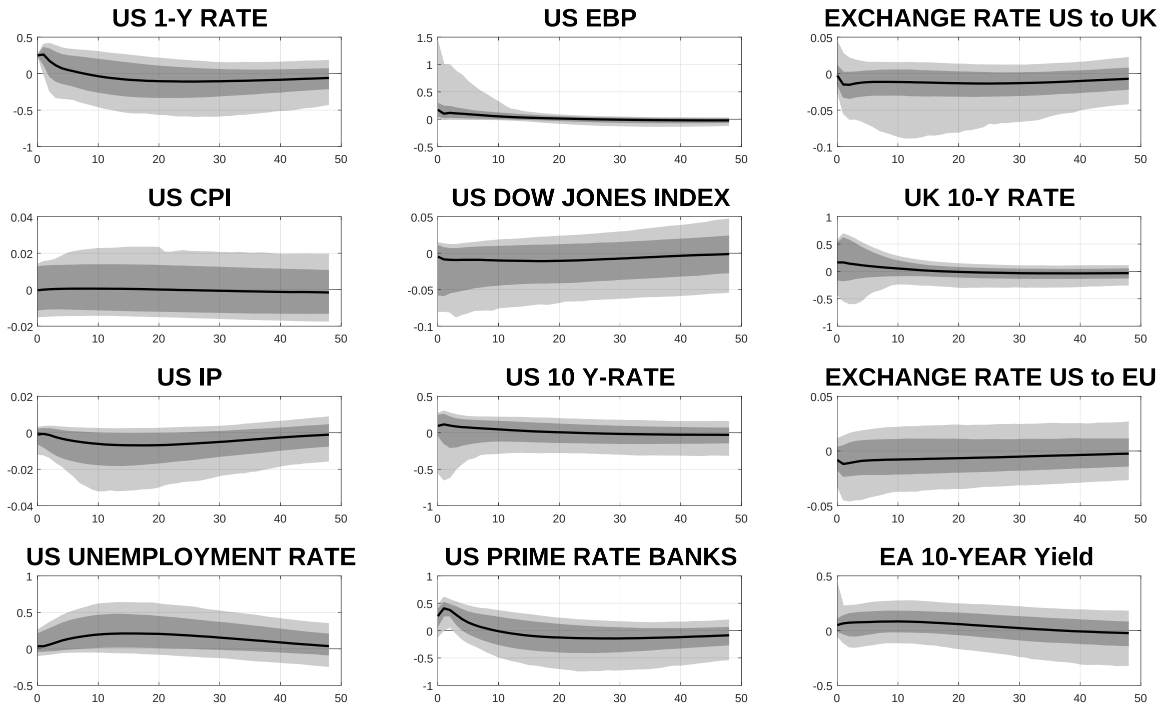
3.F.5 Results for Three-Country VAR Models with Identified US Monetary Policy Shock

Figure 3.24: Responses of foreign stock prices, exchange rates and US variables to a contractionary US monetary policy shock



Note: US monetary policy surprise series of Gertler and Karadi (2015). Responses of US, UK and EA variables to a 25 basis point increase in the US treasury 1-year rate. A sign restriction is set on the response of the excess bond premium on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

Figure 3.25: Responses of foreign interest rates, exchange rates and US variables to a contractionary US monetary policy shock



Note: US monetary policy surprise series of Gertler and Karadi (2015). Responses of US, UK and EA variables to a 25 basis point increase in the US treasury 1-year rate. A sign restriction is set on the response of the excess bond premium on the monetary policy shock. The response it set to be positive. The gray shaded areas is the 90% Bayesian credible set, the light gray shaded are the 68%. VAR model data from Jan 1973 to June 2017. Proxy data from Jan 1990 to June 2012.

Bibliography

Ando, T. and J. Bai (2016): Panel Data Models with Grouped Factor Structure Under Unknown Group Membership, *Journal of Applied Econometrics*, 31(1), 163–191.

Attinasi, M.G. and L. Metelli (2017): Is fiscal consolidation self-defeating? A panel-VAR analysis for the Euro area countries, *Journal of International Money and Finance*, 74, 147–164.

Banbura, M., D. Giannone and L. Reichlin (2010): Large Bayesian vector auto regressions, *Journal of Applied Econometrics*, 25(1), 71–92.

Banerjee, O., L. El Ghaoui and A. d'Aspremont (2008): Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data, *Journal of Machine Learning Research*, 9, 485–516.

Basu, S. and G. Michailidis (2015): Regularized estimation in sparse high-dimensional time series models, *The Annals of Statistics*, 43(4), 1535–1567.

Bauer, M. and C. Neely (2014): International channels of the Fed's unconventional monetary policy, *Journal of International Money and Finance*, 44, 24–46.

Belloni, A. and V. Chernozhukov (2013): Least squares after model selection in high-dimensional sparse models, *Bernoulli*, 19(2), 521–547.

Bergmeir, C., R.J. Hyndman and B. Koo (2018): A Note on the Validity of Cross-validation for Evaluating Autoregressive Time Series Prediction, *Computational Statistics & Data Analysis*, 120, 70–83.

Bernanke, B.S., J. Boivin and P. Eliasz (2005): Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach, *The Quarterly Journal of Economics*, 120(1), 387–422.

Billio, M., R. Casarin, F. Ravazzolo and H.K. van Dijk (2016): Interconnections Between Eurozone and US Booms and Busts Using a Bayesian Panel Markov-Switching VAR Model, *Journal of Applied Econometrics*, 31(7), 1352–1370.

Billio, M., R. Casarin and L. Rossini (2016): Bayesian nonparametric sparse seemingly unrelated regression model (SUR), Working Papers No. 2016:20, Department of Economics, University of Venice "Ca' Foscari".

Bjørnland, H.C., F. Ravazzolo and L.A. Thorsrud (2017): Forecasting GDP with global components: This time is different, *International Journal of Forecasting*, 33(1), 153–173.

Bluwstein, K. and F. Canova (2016): Beggar-Thy-Neighbor? The International Effects of ECB Unconventional Monetary Policy Measures, *International Journal of Central Banking*, 12(3), 69–120.

Boeckx, J., M. Dossche and G. Peersman (2017): Effectiveness and Transmission of the ECB's Balance Sheet Policies, *International Journal of Central Banking*, 13(1), 297–333.

Breitung, J. and C. Roling (2015): Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach, *Journal of Forecasting*, 34(7), 588–603.

Bühlmann, P. and S. van de Geer (2011): *Statistics for High-Dimensional Data. Methods, Theory and Applications*: Springer.

Burriel, P. and A. Galesi (2018): Uncovering the heterogeneous effects of ECB unconventional monetary policies across euro area countries, *European Economic Review*, 101, 210–229.

Caldara, D. and E. Herbst (2016): Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs, No. 2016-049, Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System.

Canova, F. (2005): The transmission of US shocks to Latin America, *Journal of Applied Econometrics*, 20(2), 229–251.

Canova, F. and M. Ciccarelli (2004): Forecasting and turning point predictions in a Bayesian panel VAR model, *Journal of Econometrics*, 120(2), 327–359.

Canova, F. and M. Ciccarelli (2009): Estimating Multicountry VAR Models, *International Economic Review*, 50(3), 929–959.

Canova, F. and M. Ciccarelli (2012): ClubMed? Cyclical fluctuations in the Mediterranean basin, *Journal of International Economics*, 88(1), 162–175.

Canova, F. and M. Ciccarelli (2013): Panel vector autoregressive models: a survey, Working Paper Series No. 1507, European Central Bank.

Cesa-Bianchi, A., G. Thwaites and A. Vicondoa (2016): Monetary Policy Transmission in an Open Economy: New Data and Evidence from the United Kingdom, Discussion Papers No. 1612, Centre for Macroeconomics (CFM).

Chen, Q., A. Filardo, D. He and F. Zhu (2016): Financial crisis, US unconventional monetary policy and international spillovers, *Journal of International Money and Finance*, 67, 62–81.

Ciccarelli, M., A. Maddaloni and J.L. Peydro (2013): Heterogeneous transmission mechanism: monetary policy and financial fragility in the euro area, Working Paper Series No. 1527, European Central Bank.

Ciccarelli, M. and B. Mojon (2010): Global Inflation, *The Review of Economics and Statistics*, 92(3), 524–535.

Ciccarelli, M., E. Ortega and M.T. Valderrama (2016): Commonalities and cross-country spillovers in macroeconomic-financial linkages, *The B.E. Journal of Macroeconomics*, 16(1), 231–275.

Comunale, M. (2017): A panel VAR analysis of macro-financial imbalances in the EU, Working Paper Series No. 2026, European Central Bank.

Davis, R., P. Zang and T. Zheng (2016): Sparse Vector Autoregressive Modeling, *Journal of Computational and Graphical Statistics*, 25(4), 1–53.

Dees, S., F. di Mauro, L.V. Smith and M.H. Pesaran (2007): Exploring the international linkages of the euro area: a global VAR analysis, *Journal of Applied Econometrics*, 22(1), 1–38.

Doan, T., R. Litterman and C. Sims (1984): Forecasting and conditional projection using realistic prior distributions, *Econometric Reviews*, 3(1), 1–100.

Dovern, J., M. Feldkircher and F. Huber (2016): Does joint modelling of the world economy pay off? Evaluating global forecasts from a Bayesian GVAR, *Journal of Economic Dynamics and Control*, 70, 86–100.

Faust, J., J. Rogers, E. Swanson and J. Wright (2003): Identifying the Effects of Monetary Policy Shocks on Exchange Rates Using High Frequency Data, *Journal of the European Economic Association*, 1(5), 1031–1057.

Faust, J., E. Swanson and W.J. H (2004): Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?, *The B.E. Journal of Macroeconomics*, 4(1), 1–31.

Fratzscher, M., M.L. Duca and R. Straub (2016): ECB Unconventional Monetary Policy: Market Impact and International Spillovers, *IMF Economic Review*, 64(1), 36–74.

Fratzscher, M., M.L. Duca and R. Straub (2018): On the International Spillovers of US Quantitative Easing, *The Economic Journal*, 128(608), 330–377.

Friedman, J., T. Hastie, H. Höfling and R. Tibshirani (2007): Pathwise coordinate optimization, *Ann. Appl. Stat.*, 1(2), 302–332.

Friedman, J., T. Hastie and R. Tibshirani (2008): Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, 9(3), 432–441.

Friedman, J., T. Hastie and R. Tibshirani (2010): Regularization Paths for Generalized Linear Models via Coordinate Descent, *Journal of statistical software*, 33(1), 1–22.

Gambacorta, L., B. Hofmann and G. Peersman (2014): The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis, *Journal of Money, Credit and Banking*, 46(4), 615–642.

Garratt, A., K. Lee and K. Shields (2016): Forecasting global recessions in a GVAR model of actual and expected output, *International Journal of Forecasting*, 32(2), 374–390.

Gefang, D. (2014): Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage, *International Journal of Forecasting*, 30(1), 1–11.

George, E.I. and R.E. McCulloch (1993): Variable Selection via Gibbs Sampling, *Journal of the American Statistical Association*, 88(423), 881–889.

George, E.I., D. Sun and S. Ni (2008): Bayesian stochastic search for VAR model restrictions, *Journal of Econometrics*, 142(1), 553–580.

Georgiadis, G. (2015): Examining asymmetries in the transmission of monetary policy in the euro area: Evidence from a mixed cross-section global VAR model, *European Economic Review*, 75, 195–215.

Georgiadis, G. (2017): To bi, or not to bi? Differences between spillover estimates from bilateral and multilateral multi-country models, *Journal of International Economics*, 107, 1–18.

Gerko, E. and H. Rey (2017): Monetary Policy in the Capitals of Capital, *Journal of the European Economic Association*, 15(4), 721–745.

Gertler, M. and P. Karadi (2015): Monetary Policy Surprises, Credit Costs, and Economic Activity, *American Economic Journal: Macroeconomics*, 7(1), 44–76.

Giannone, D., M. Lenza and G.E. Primiceri (2015): Prior Selection for Vector Autoregressions, *The Review of Economics and Statistics*, 97(2), 436–451.

Glick, R. and S. Leduc (2012): Central bank announcements of asset purchases and the impact on global financial and commodity markets, *Journal of International Money and Finance*, 31(8), 2078–2101.

Gnimassoun, B. and V. Mignon (2016): How Do Macroeconomic Imbalances Interact? Evidence from a Panel VAR Analysis., *Macroeconomic Dynamics*, 20(7), 1717–1741.

Greenwood-Nimmo, M., V.H. Nguyen and Y. Shin (2012): Probabilistic forecasting of output growth, inflation and the balance of trade in a GVAR framework, *Journal of Applied Econometrics*, 27(4), 554–573.

Gürkaynak, R.S., B. Sack and E. Swanson (2005): Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements, *International Journal of Central Banking*, 1(1), 55–93.

Hachula, M., M. Piffer and M. Rieth (2016): Unconventional Monetary Policy, Fiscal Side Effects and Euro Area (Im)balances, Discussion Papers of DIW Berlin No. 1596, DIW Berlin, German Institute for Economic Research.

Hansen, B.E. (2016): The Risk of James-Stein and Lasso Shrinkage, *Econometric Reviews*, 35(8-10), 1456–1470.

Hastie, T., R. Tibshirani and M. Wainwright (2015): *Statistical Learning with Sparsity. The Lasso and Generalizations*, Vol. 143 of Monographs on Statistics and Applied Probability.

Hausman, J. and J. Wongswan (2011): Global asset prices and FOMC announcements, *Journal of International Money and Finance*, 30(3), 547–571.

Hsu, N.J., H.L. Hung and Y.M. Chang (2008): Subset selection for vector autoregressive processes using Lasso, *Computational Statistics & Data Analysis*, 52(7), 3645–3657.

Huber, F., M. Feldkircher and J.C. Cuaresma (2016): Forecasting with Global Vector Autoregressive Models: A Bayesian Approach, *Journal of Applied Econometrics*, 31(7), 1371–1391.

Jarociński, M. and P. Karadi (2017): Central Bank Information Shocks, 2017 Meeting Papers No. 1193, Society for Economic Dynamics.

Jarociński, M. and B. Maćkowiak (2017): Granger Causal Priority and Choice of Variables in Vector Autoregressions, *The Review of Economics and Statistics*, 99(2), 319–329.

Jochmann, M., G. Koop and R.W. Strachan (2010): Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks, *International Journal of Forecasting*, 26(2), 326–347.

Karlsson, S. (2013). *Handbook of Economic Forecasting*, Vol. 2, Chapter Forecasting with Bayesian vector autoregessions, pp. 791–897. Amsterdam: Elsevier.

Kascha, C. and C. Trenkler (2015): Forecasting VARs, model selection, and shrinkage, Working Paper ECON 15-07, University of Mannheim.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*, Chapter 16. Structural VAR Analysis in a Data-Rich Environment, pp. 535–574: Cambridge University Press.

Kim, S. (2001): International transmission of U.S. monetary policy shocks: Evidence from VAR's, *Journal of Monetary Economics*, 48(2), 339–372.

Kim, S. and N. Roubini (2000): Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach, *Journal of Monetary Economics*, 45(3), 561–586.

Kock, A.B. and L. Callot (2015): Oracle inequalities for high dimensional vector autoregressions, *Journal of Econometrics*, 186(2), 325–344.

Koop, G. (2013): Forecasting with Medium and Large Bayesian VARs, *Journal of Applied Econometrics*, 28(2), 177–203.

Koop, G. and D. Korobilis (2010): Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, *Foundations and Trends(R) in Econometrics*, 3(4), 267–358.

Koop, G. and D. Korobilis (2015a): Forecasting With High Dimensional Panel VARs, Working Papers No. 2015:25, Business School - Economics, University of Glasgow.

Koop, G. and D. Korobilis (2015b): Model uncertainty in Panel Vector Autoregressive models, *European Economic Review*, 81, 115–131.

Koop, G., R. Strachan, H. van Dijk and M. Villani (2005). *Palgrave Handbook of Econometrics*, Vol. Volume 1: Econometric Theory, Chapter Bayesian approaches to cointegration. Houndmills: Palgrave Macmillan.

Korobilis, D. (2008): Forecasting in Vector Autoregressions with Many Predictors, *Advances in Econometrics, 23: Bayesian Macroeconometrics*, 409–431.

Korobilis, D. (2013): VAR Forecasting Using Bayesian Variable Selection, *Journal of Applied Econometrics*, 28(2), 204–230.

Korobilis, D. (2016): Prior selection for panel vector autoregressions, *Computational Statistics and Data Analysis*, 101, 110–120.

Kucharukov, O.B., P. Claeys and B. Vasicek (2016): Spillover of the ECB's monetary policy outside the euro area: How different is conventional from unconventional policy?, *Journal of Policy Modeling*, 38(2), 199–225.

Kuttner, K.N. (2001): Monetary policy surprises and interest rates: Evidence from the Fed funds futures market, *Journal of Monetary Economics*, 47(3), 523–544.

Kyung, M., J. Gill, M. Ghosh and G. Casella (2010): Penalized regression, standard errors, and Bayesian lassos, *Bayesian Analysis*, 5(2), 369–411.

Lee, W. and Y. Liu (2012): Simultaneous multiple response regression and inverse covariance matrix estimation via penalized Gaussian maximum likelihood, *Journal of Multivariate Analysis*, 111, 241–255.

Liao, Z. and P.C.B. Phillips (2015): Automated Estimation of Vector Error Correction Models, *Econometric Theory*, 31(3), 581–646.

Litterman, R. (1986): Forecasting with Bayesian Vector Autoregressions-Five Years of Experience, *Journal of Business & Economic Statistics*, 4(1), 25–38.

Love, I. and L. Zicchino (2006): Financial development and dynamic investment behavior: Evidence from panel VAR, *The Quarterly Review of Economics and Finance*, 46(2), 190–210.

Maćkowiak, B. (2007): External shocks, U.S. monetary policy and macroeconomic fluctuations in emerging markets, *Journal of Monetary Economics*, 54(8), 2512–2520.

Marcellino, M., J.H. Stock and M.W. Watson (2006): A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of Econometrics*, 135(1-2), 499–526.

Mertens, K. and M.O. Ravn (2013): The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States, *American Economic Review*, 103(4), 1212–47.

Miranda-Agrippino, S. (2016): Unsurprising shocks: information, premia, and the monetary transmission, Bank of England working papers No. 626, Bank of England.

Nakamura, E. and J. Steinsson (forthcoming): High frequency identification of monetary non-neutrality: The information effect, *Quarterly Journal of Economics*.

Neely, C.J. (2015): Unconventional monetary policy had large international effects, *Journal of Banking & Finance*, 52, 101–111.

Ngueyep, R. and N. Serban (2015): Large-Vector Autoregression for Multilayer Spatially Correlated Time Series, *Technometrics*, 57(2), 207–216.

Nicholson, W.B., D.S. Matteson and J. Bien (2016): High Dimensional Forecasting via Interpretable Vector Autoregression, *ArXiv e-prints*.

Nicholson, W.B., D.S. Matteson and J. Bien (2017): VARX-L: Structured regularization for large vector autoregressions with exogenous variables, *International Journal of Forecasting*, 33(3), 627–651.

Obstfeld, M. and K. Rogoff (2002): Global Implications of Self-Oriented National Monetary Rules, *The Quarterly Journal of Economics*, 117(2), 503–535.

Park, T. and G. Casella (2008): The Bayesian Lasso, *Journal of the American Statistical Association*, 103(482), 681–686.

Pérez, F. (2015): Comparing the Transmission of Monetary Policy Shocks in Latin America: A Hierarchical Panel VAR, Working Papers No. 2015-015, Banco Central de Reserva del Peru.

Pesaran, M.H., T. Schuermann and L.V. Smith (2009): Forecasting economic and financial variables with global VARs, *International Journal of Forecasting*, 25(4), 642–675.

Ren, Y., Z. Xiao and X. Zhang (2013): Two-step adaptive model selection for vector autoregressive processes, *Journal of Multivariate Analysis*, 116, 349–364.

Ren, Y. and X. Zhang (2010): Subset selection for vector autoregressive processes via adaptive Lasso, *Statistics & Probability Letters*, 80(23-24), 1705–1712.

Rogers, J.H., C. Scotti and J.H. Wright (2014): Evaluating asset-market effects of unconventional monetary policy: a multi-country review, *Economic Policy*, 29(80), 749–799.

Rogers, J.H., C. Scotti and J.H. Wright (2017): Unconventional Monetary Policy and International Risk Premia, International Finance Discussion Papers No. 1172, Board of Governors of the Federal Reserve System (U.S.).

Schnücker, A. (2016): Restrictions search for panel VARs, DIW Discussion Papers No. 1612, Berlin.

Song, S. and P.J. Bickel (2011): Large Vector Auto Regressions, *ArXiv e-prints*.

Stock, J.H. and M.W. Watson (2012): Disentangling the Channels of the 2007-09 Recession, *Brookings Papers on Economic Activity*, 43(1), 81–156.

Su, L., Z. Shi and P.C.B. Phillips (2016): Identifying Latent Structures in Panel Data, *Econometrica*, 84(6), 2215–2264.

Taylor, J.B. (2013): International monetary coordination and the great deviation, *Journal of Policy Modeling*, 35(3), 463–472.

Tibshirani, R. (1996): Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Socienty, Series B: Statistical Methodology*, 58(1), 267–288.

Wieladek, T. (2016): The varying coefficient Bayesian panel VAR model, Bank of England working papers No. 578, Bank of England.

Yuan, M. and Y. Lin (2006): Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zou, H. (2006): The Adaptive Lasso and Its Oracle Properties, *Journal of the American Statistical Association*, 101(476), 1418–1429.

Zou, H. and T. Hastie (2005): Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorgelegte Dissertation auf Grundlage der angegebenen Quellen und Hilfsmittel selbstständig verfasst habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht. Die vorgelegte Dissertation hat weder in der gleichen noch einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule zum Promotionsverfahren vorgelegen.

Annika Schnücker
Berlin, den 22. Mai 2018

Liste verwendeter Hilfsmittel

- Matlab R2016a
  - Optimization Toolbox
  - Statistics Toolbox
  - Parallel Computing Toolbox
- RStudio 1.1.383 basierend auf R 3.3.0
- Microsoft Excel
- LaTeX
- Siehe auch Literatur- und Quellenangaben