THE CONSISTENCY OF QUANTILE REGRESSION IN LINEAR MIXED MODELS

Dissertation submitted in fulfillment of the requirements for the degree
Dr. rer. pol. (Doctor of Economic Sciences)
School of Business & Economics
Freie Universität Berlin

by

DIPL.-MATH. BEATE WEIDENHAMMER

in

Berlin, June 2016

SUPERVISORS:
Prof. Dr. Timo Schmid
Prof. Dr. Nikos Tzavidis

LOCATION:
Berlin

DATE OF DISPUTATION:
May 30th, 2017

DECLARATION

I certify that this work contains no material, which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

Berlin, March 2019

Dipl.-Math. Beate Weidenhammer,
March 28, 2019

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 PREFACE

1.1 MOTIVATION

Quantiles of a random variable are parameters of different ways of interpretation. They are location and scale parameters at the same time. For a given $\tau \in (0, 1)$ the τ-quantile is the value of a random variable, where $\tau \cdot 100\%$ of its values lie below. Thus quantiles for τ close to 0.5 give the location of the random variable and quantiles for τ close to zero or one give an idea of the spread in the random variable. Furthermore parameters like the 0.5-quantile or median are robust while the mean is not. Therefore the median outperforms the mean in this regard. A set of a few quantiles can already give a good overview of the distribution of data. A boxplot, for example, needs the median, the lower and the upper quartiles (the 0.25- and the 0.75-quantiles). By adding the lowest and highest observations a clear picture is drawn.

In this thesis I am going to prove the consistency of the quantile estimator in linear mixed models. The idea of quantile estimation in linear models was firstly introduced by Koenker and Bassett [1978]. Based on this idea several further properties and extensions were developed. So is the equivalence to an *asymmetric Laplacian* linear model one main property of the quantile linear model. It was on the median model introduced by Jung [1996], where he used a *Laplacian* linear model. Koenker and Machado [1999] extended this idea to other quantiles using the *asymmetric Laplace distribution*. The idea shifts the quantile regression as a minimising problem with a least absolute deviation approach to a *maximum likelihood* estimation. This field is well developed and it can be shown that the estimator is *asymptotically normal* distributed, which implies the consistency. Latter is a wishful feature of estimators, because it means the convergence to the real value of interest with increasing sample size. At the same time the variance of the estimator, which is again a random variable, decreases. As a result the estimation is asymptotically unbiased and the *mean squared error* (*MSE*) converges to zero.

The quantile regression was extended to regression of count data using a method called jittering by Machado and Santos Silva [2005]. Not only that the quantiles of integers are integers, they are also less prone to overdispersion. The misspecification of the variance of the data is one of the main problem of the classical Poisson model approach. The median as a robust location parameter may be a solution to this. On top it gives a good approach on the estimation of other quantiles of counts, which is proven to be consistent in Machado and

Santos Silva [2005].
In longitudinal data or in the field of *Small Area Estimation* (SAE) the linear model is extended to a linear mixed model by adding a random effect. The random effect is added such that we avoid problems with the independence of the data points which is assumed on the error terms. For example in longitudinal data the data points of one object or person over time would easily violate the assumption of independent observations in a linear model. The same violation appears in SAE with observations from the same area. In linear mixed models the random effect is the same for the same area or object in longitudinal data. Thus the error terms will still be independent. Nevertheless in these models mean estimators are well investigated. The idea of quantile estimation in linear mixed models, however, was introduced by Geraci and Bottai [2007] and Geraci and Bottai [2014]. They used the equivalence of the linear quantile mixed model to an *asymmetric Laplacian* linear mixed model and employed numerical approaches for the *maximum likelihood* estimation. This is necessary because a closed form solution of the maximisation of the log-likelihood density is not existent, due to a rather complicated appearance of the density. In Geraci and Bottai [2007] the authors use an *EM algorithm* (cf. McLachlan and Krishnan [2008]), while they employ a *Gaussian quadrature* (cf. Pinheiro and Chao [2006]) in Geraci and Bottai [2014]. The latter was implemented in the open source software R in a package called lqmm (cf. Geraci [2016]). As the mean estimator the approach is a two-stage method. There the unknown parameters from the density of the observations are estimated by a *maximum likelihood* estimation in a first step. A predictor for the random effect is derived from the *maximum likelihood* estimator in a second step.
So far an asymptotic theory on the quantile estimation in linear mixed models has not been developed. For the mean estimation in linear mixed models Miller [1977] and Pinheiro [1994] showed the *asymptotic normality* (cf. Chapter 9.3 of van der Vaart [2007]) of the *maximum likelihood* estimator in the first step. For this reason they applied a theorem proved by Weiss [1971] and Weiss [1973]. This theorem, the so called *Weiss' Theorem*, is a specialisation of the *Glivenko-Cantelli Theorem* (cf. Glivenko [1933] and Cantelli [1933]) in the case of non-independent observations. In the mean case the individual error terms were assumed to be *normal* and so is the random effect. The consequential *normality* for the observation yields the *maximum likelihood* estimator in the the first step. The assumptions of the *Weiss' Theorem* are convergences of the second derivatives of the log-likelihood density, which represent a *Fisher Information* kind matrix. Searle et al. [1992] deduced the derivatives leading into a straightforward application in the mean case.
For the quantile estimation in linear mixed models the *Weiss' Theorem* is also applicable. In order to do so, its assumptions must be proven,

which turns out to be elaborate. This is the main achievement of this thesis while there are some further assumptions added. In the linear quantile mixed model the assumption of *asymmetric Laplacian* individual error terms has no direct implication on the distribution of the observation. Thus the log-likelihood density is displayed as an integral and the derivatives cannot be calculated directly. This impacts the application of the *Weiss' Theorem* and, hence, the proof of the *asymptotic normality* of the *maximum likelihood* estimator in the first step.
In the second step of both the quantile and the mean estimation the random effect is predicted. The consistency of the predictor in the quantile estimation is shown in this thesis. The *asymptotic normality* of the *maximum likelihood* estimator of the first step implies its consistency. Eventually the consistency of the quantile estimator in linear mixed models under the assumptions made is established.
The property of an estimator or predictor to be consistent is of important role in statistics. In words it means the convergence of the quantile estimator to the true quantile in linear mixed models, whenever the sample size increases. It implies the decline of both the variance and the bias of the quantile estimator. Since the sample size in linear mixed models is dependent on the number of groups and the within group sample size, both numbers must increase in order to accomplish consistency. Pinheiro [1994] assumed this property for his demonstration of the consistency in the mean estimation and so will I undertake this in the quantile estimation. Furthermore, compared to Pinheiro [1994], I added an additional assumption which is rather of constructional character. However they are only additional assumptions next to classical regularity conditions in *Glivenko-Cantelli* kind statements.

1.2 OUTLINE

In this thesis I begin with an introduction to quantile estimation in general. Chapter 2 starts with quantiles and their estimation in Section 2.1. There I give a convention on the definition of a quantile. For a given $\tau \in (0, 1)$ a τ-quantile is the value of a random variable, where $\tau \cdot 100\%$ of its values lie below. It happens that this is not one-to-one and onto, which requires the convention which is also employed in Koenker [2005]. Furthermore I discuss the loss function of the estimation and compare it with the classical squared loss in mean estimation.
In the following Section 2.2 I shortly illustrate the mean estimation in linear models, which leads to the introduction of the linear quantile model in Section 2.3. Within this Section I show the parameter estimation in the model leading to a quantile estimator in linear models. This is followed, in Section 2.4, by the proof of the equivalence

of the linear quantile model to a linear model with *asymmetric Laplacian* error terms. In Section 2.4.1 the *asymmetric Laplace distribution* is discussed. The equivalence of the two models implies the possibility of two different approaches in estimation: a minimising problem and a *maximum likelihood* estimation. These two methods carry the same properties and in Section 2.5 I discuss the consistency of the quantile estimation, which follows from both of them. In Section 2.6 I then give a short overview on further research questions and their solutions for the linear quantile model.
The Chapter is ended by two applications and extensions of the model in Section 2.7. There, in Section 2.7.1, I discuss the employment of link functions in quantile regression and the perpetuation of the consistency for the generalised linear quantile models. In the last Section 2.7.2 the quantile estimation is applied on count data. There an overview of the different steps of the estimation is given and the consistency of the quantile estimator for count data is shown.
Chapter 3 builds the main part of this thesis. The linear quantile mixed model is discussed in terms of estimation and its consistency. Before I delve into the quantile estimation in linear mixed models, I discuss the mean estimation in Section 3.1. The mean model always operates as a starting point for quantile estimation.
In Section 3.2 I then introduce the linear quantile mixed model and its equivalence to a linear mixed model with *asymmetric Laplacian* error terms. The approach is similar to the approach in linear models with no random effect. However in the mixed models the *maximum likelihood* estimation is better performable than the minimisation of a loss function approach. The linear quantile mixed model is then further discussed in Section 3.2.2. In this part I argue the dependency of the random effect on the choice of τ. In order to found this statement I show a simulation study with different error distribution scenarios.
The two steps of estimation of a quantile in a linear mixed model are shown in Section 3.3. There the *maximum likelihood* approach is followed. The consistency of this method is proven in Section 3.4. Theorem 3.1 states this main achievement of the thesis. Since the estimation is fulfilled in two steps, I also execute the proof of the consistency of it in two steps. The first step is the parameter estimation with a *maximum likelihood* approach. Since the observations are not independent of each other in a mixed model, the consistency does not just follow by the assumption of classical regularity assumptions. Thus the proof is rather extensive and is executed in detail in a separate Chapter 4. In the second step the random effect is predicted. Its consistency can be shown to follow from the consistency of the parameter estimator in the first step. A simulation study in Section 3.4.2 certifies the consistency of the quantile estimation in linear mixed models.
In the following I discuss the influence of the consistency of the quantile estimation on the *mean squared error* (*MSE*) in Section 3.5. There

it follows the convergence of the *MSE* to zero with increasing sample size.
At the end of the Chapter two applications and extensions of the linear quantile mixed model are presented. In Section 3.6.1 I show that the model can be adapted to count data, while the consistency of the estimation remains in this case. Furthermore I developed a method called *Microsimulation via Quantiles* (MvQ), which I describe in Section 3.6.2. It can be applied for the estimation of group parameters, which are beyond means. For this approach I use the natural connection between quantiles and the distribution function of a random variable. By estimating quantiles for a grid of τ I get an empirical distribution, from which each parameter of interest can be derived. This is executed via a *Monte Carlo* simulation or microsimulation.
The last Chapter 4 is the outsourced part of the consistency proof in Theorem 3.1. For the *maximum likelihood* approach in the parameter estimation I use the so-called *Weiss' Theorem* for the dependent observations. It is stated and observed in Section 4.1.1. There I discuss its assumptions made on the second derivatives of the log-likelihood density and their meaning. Furthermore I display applications of the theorem, especially for the mean estimation in linear mixed models. In preparation of the application of the theorem I then deduce the second derivatives of the log-likelihood density in Section 4.1.3. This part is rather complex, which is why it is divided into several subsections and the results are stated in lemmata.
After the preliminaries I prove the two assumptions of the *Weiss' Theorem* in Section 4.2. By application of the theorem this then yields the *asymptotic normality* of the parameter estimators. The implication of this property to consistency and the asymptotic covariance matrix is shown and discussed in Sections 4.3.1 and 4.3.2, respectively. A simulation study in Section 4.4 completes the Chapter. There the *asymptotic normality* is shown for the parameter estimation for two different quantiles and the covariance matrices are discussed in comparison with the analytic results.

2 QUANTILE REGRESSION IN LINEAR MODELS

In this Chapter I give an introduction to quantile regression in linear models. This idea is mainly based in the article Koenker and Bassett [1978] and the book Koenker [2005]. Before I start with the regression problem I will give an overview on quantiles in general in Section 2.1. There I define quantiles and give an instruction on their estimation. Before I describe the quantile estimation in regression models, I discuss the linear model and the mean estimation in Section 2.2 based on Searle [1997] and Searle et al. [1992]. The linear model provides a starting point for linear quantile models. The latter are introduced in Section 2.3 orientated on Koenker [2005]. In the following I will show again the equivalence of the linear quantile model to a linear model with *asymmetric Laplacian distributed* error terms in Section 2.4 which is a classical approach in quantile regression. Within this Section I will also give a detailed discussion on the *asymmetric Laplace distribution* mainly based on Yu and Zhang [2005]. In the end we have two ways of quantile estimation in linear models due to the two equivalent models. One is a classical minimising problem and the other one is a *maximum likelihood* approach. Section 2.5 considers the asymptotic behaviour of the estimation. It can be shown in several ways that the estimation is consistent (cf. Koenker [2005] or Pollard [1991]). In Section 2.6 I draw an overview on further research fields based on the ideas of quantile estimation. Eventually I give two possible applications of the quantile regression in linear models. In Section 2.7.1 I show that the quantile regression may also work on generalised linear models with continuous link functions without loss of the consistency in the estimation. A further application is the quantile estimation of count data in Section 2.7.2. This work is based on Machado and Santos Silva [2005] where they introduced a way of estimating integers for count quantiles and showed the consistency of this approach.

2.1 QUANTILES AND THEIR ESTIMATION

Occasionally in regression models we are interested in estimators other than the mean. So we might be keen to estimate a quantile of the dependent variable Y. Quantiles are parameters which may describe the location and the dispersion of a random variable. On top the median estimator is a robust location parameter. In mean estimation there are several approaches on robustification of the estimation based on Huber [1964] (see Section 2.3.1 for further details on robust estimation). A median is another approach of robust estimation.

Let us assume that the random variable Y follows a distribution P_Y, which is described by its distribution function

$$F_Y(y) := P_Y(Y \leqslant y).$$

The quantile function of a random variable is the inverse of the distribution function. Since a distribution function must not be strictly monotonic and thus bijective, this inverse is often not unique. This is why I introduce a convention at this point which follows the definition in Koenker [2005]. In the case of non-bijectivity, the quantile is defined as the smallest value of the set of values, where the distribution function maps to an output. We define it as follows

$$Q_Y(\tau) = F_Y^{-1}(\tau) := \inf\{y | F_Y(y) \geqslant \tau\} \tag{2.1}$$

for $0 \leqslant \tau \leqslant 1$. In words, a τ-quantile $Q_Y(\tau)$ is the smallest value of the domain of Y, such that at least $\tau \cdot 100\%$ of the values lie below this value. In conclusion it is the value such that at most $(1-\tau) \cdot 100\%$ of the values lie above. This leads to the before mentioned two functions of quantiles as parameters of location and of scale. The median is defined as the 50%-quantile and is a well known location parameter

$$median(Y) := Q_Y(0.5).$$

On the other hand quantiles for small and large values of τ give an impression of the scale. The interquartile range, as an example, is the difference between the lower quartile $Q_Y(0.25)$ and the upper quartile $Q_Y(0.75)$

$$IQR(Y) := Q_Y(0.75) - Q_Y(0.25).$$

Whenever we are interested in estimating a parameter θ of the distribution of a random variable a risk $R(\theta)$ is needed. It is defined as the expectation of a given loss function L under the measure P_Y

$$R(\theta) := E_Y[L(\theta, Y)].$$

Generally the risk depends on a loss function $L(\theta, Y)$ and the estimator $\hat{\theta}$ is given as the minimiser of it:

$$\hat{\theta} := \arg\min_{\theta} E_Y[L(\theta, Y)].$$

For example (cf. Searle [1971]) in mean estimation we find the estimator $\theta = E_Y[Y]$ by minimising the expectation of the squared differences

$$\min_{\theta} E_Y\left[(Y-\theta)^2\right] = \min_{\theta} \int (y-\theta)^2 dP_Y(y)$$

leading to

$$E_Y[Y] = \arg\min_{\theta} E_Y\left[(Y-\theta)^2\right].$$

The loss function is the squared difference and the risk is the squared 2-norm with respect to the measure P_Y

$$\|a\|_{2,P_Y}^2 := \int a^2 dP_Y(a).$$

In application for a sample $Y_1, Y_2, \dots, Y_n \overset{iid}{\sim} P_Y$ the mean estimator is given by

$$\bar{Y} = \arg\min_{\theta} \sum_{i=1}^{n} (Y_i - \theta)^2.$$

Thus the mean estimation is based on the *Euclidean distance*

$$\|a\|_2 := \sqrt{\sum_{i=1}^{n} a_i^2}$$

as the empirical risk measure.
In quantile estimation with the parameter of interest $\theta = Q_Y(\tau)$ for a given $\tau \in (0, 1)$ Fox and Rubin [1964] showed that the loss function is based on the distance measure

$$\rho_\tau(a) := a\left(\tau - \mathbf{1}_{\{a<0\}}\right) \tag{2.2}$$

where $\mathbf{1}_{\{.\}}$ stands for the indicator function which is one if the condition in the footnote is fulfilled and zero otherwise. The distance function may be rewritten as follows

$$\begin{aligned} \rho_\tau(a) &:= a\left(\tau - \mathbf{1}_{\{a<0\}}\right) \\ &= |a|\left(\tau\mathbf{1}_{\{a\geqslant 0\}} + (1-\tau)\right). \end{aligned}$$

This leads together with (2.2) to the risk function as the expected loss

$$\begin{aligned} R(\theta) &= E_Y\left[L(\theta, Y)\right] \\ &= E_Y\left[\rho_\tau(Y-\theta)\right] \\ &= \int \rho_\tau(y-\theta) dP_Y(y) \\ &= \int (y-\theta)\cdot(\tau - \mathbf{1}_{\{y-\theta<0\}}) dP_Y(y) \\ &= \int (y-\theta)\cdot(\tau - \mathbf{1}_{\{y<\theta\}}) dP_Y(y) \\ &= \tau \int_{-\infty}^{\theta} |y-\theta|\, dP_Y(y) + (1-\tau)\cdot \int_{\theta}^{\infty} |y-\theta|\, dP_Y(y). \end{aligned}$$

Differentiating the risk function with respect to θ leads to

$$\begin{aligned}
0 &\stackrel{!}{=} \frac{\partial}{\partial\theta}\left(\tau\int_{-\infty}^{\theta}|y-\theta|\,dP_Y(y)+(1-\tau)\cdot\int_{\theta}^{\infty}|y-\theta|\,dP_Y(y)\right)\\
&= -\tau\int_{-\infty}^{\theta}dP_Y(y)+(1-\tau)\cdot\int_{\theta}^{\infty}dP_Y(y)\\
&= \int_{\theta}^{\infty}dP_Y(y)-\tau\int_{\theta}^{\infty}dP_Y(y)\\
&= F_Y(\theta)-\tau.
\end{aligned}$$

Note that the latter derivation was only executed such that the reader has a better understanding of the link between the distance measure ρ_τ defined in (2.2) and quantile estimation. In a different manner this link has already been derived in e.g. Koenker [2005]. Since by definition any distribution function F_Y is monotone any element of $\{y|F_Y(y)=\tau\}$ minimises the risk. Especially the quantile as defined in (2.1) is a solution with the convention that F_Y is left-continuous.
Applied to a sample $Y_1, Y_2, \ldots, Y_n \stackrel{iid}{\sim} P_Y$ the quantile estimator $\hat{Q}_Y(\tau)$ for a given $\tau \in (0,1)$ is given by

$$\hat{Q}_Y(\tau) := \arg\min_{\theta}\sum_{i=1}^{n}\rho_\tau(Y_i-\theta).$$

In detail we get

$$\begin{aligned}
\hat{Q}_Y(\tau) &= \arg\min_{\theta}\sum_{i=1}^{n}\rho_\tau(Y_i-\theta)\\
&= \arg\min_{\theta}\left((1-\tau)\sum_{\{i=1,2,\ldots,n|Y_i<\theta\}}|Y_i-\theta|\right.\\
&\qquad\left.+\tau\sum_{\{i=1,2,\ldots,n|Y_i\geqslant\theta\}}|Y_i-\theta|\right).
\end{aligned}$$

Thus the quantile estimator $\hat{Q}_Y(\tau)$ is the point where the absolute distance of all observations below are weighted with $1-\tau$ and the ones above are weighted with τ.
Before I discuss the quantile regression in linear models I give a small introduction of linear models and mean estimation in the following Section. This always acts as a starting point for the quantile approach.

2.2 THE LINEAR MODEL

The introduction of the linear model is based on Searle et al. [1992] and Searle [1997]. Its purpose is the preparation for the linear quantile model in the following Section 2.3.

Suppose we have n observations Y_i, $i = 1, 2, \ldots, n$ which are each linearly dependent on a regressor x_i which is a p-dimensional vector $x_i = (x_{i,1}, \ldots, x_{i,p})^T$. In general we consider a regression model with an intercept leading to $x_{i,1} = 1 \quad \forall i = 1, 2, \ldots, n$. In the following we assume an intercept and thus the regressor is given as $x_i = (1, x_{i,2}, \ldots, x_{i,p})^T$. The linear regression model is stated as followed

$$Y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{2.3}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is an unknown p-dimensional parameter vector and ε_i, $i = 1, 2, \ldots, n$ is the error term following a centred distribution

$$\varepsilon_i \overset{iid}{\sim} P_\varepsilon, \quad i = 1, 2, \ldots, n \quad \text{with} \quad E[\varepsilon_i] = 0$$

and finite variance (see Searle [1971] for details). So far this model is flexible in terms of concrete assumptions on the distribution of ε. As a generalisation of this model we can write it in matrix form as follows

$$Y = \mathbf{X}\beta + \varepsilon$$

where $Y = (Y_1, Y_2, \ldots, Y_n)^T$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T$ are n-dimensional random vectors and $\mathbf{X}$ is the design matrix with the auxiliary vectors x_i as the i^{th} row

$$\mathbf{X} := \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{pmatrix}.$$

For the conditional mean estimation β is estimated by solving

$$\min_{\beta \in \mathbb{R}^p} \|Y - \mathbf{X}\beta\|_2^2 = \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - x_i^T \beta)^2$$

leading to

$$\hat{\beta}(Y|\mathbf{X}) = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - x_i^T \beta)^2 \tag{2.4}$$

and the conditional mean estimator of Y given $\mathbf{X}$

$$\hat{Y} = \mathbf{X}\hat{\beta}. \tag{2.5}$$

2.3 THE LINEAR QUANTILE MODEL

As illustrated in Koenker [2005] and analogously as in the presented mean estimation, the quantile estimation for a fixed $\tau \in (0,1)$ is fulfilled by employing ρ_τ as defined in (2.2) leading to

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - x_i^T \beta). \tag{2.6}$$

Now the minimiser

$$\hat{\beta}_\tau(Y|\mathbf{X}) = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - x_i^T \beta) \tag{2.7}$$

is dependent on τ and represents a conditional quantile estimator

$$\hat{Q}_{Y_i|x_i}(\tau) = x_i^T \hat{\beta}_\tau, \quad i = 1, 2, \ldots, n. \tag{2.8}$$

Let us go back to the linear model as introduced in (2.3) with independently identical distributed errors. If we knew the distribution of the error term ε, characterised by its distribution function F_ε, we are able to state the conditional quantile of Y given x for a given $\tau \in (0,1)$ as follows

$$Q_{Y_i|x_i}(\tau) = x_i^T \beta + F_\varepsilon^{-1}(\tau), \quad i = 1, 2, \ldots, n.$$

In this case the quantile estimator $\hat{\beta}_\tau$ is just a vertical displacement of the mean estimator $\hat{\beta}$ as defined in (2.4)

$$\hat{\beta}_\tau = (\hat{\beta}_1 + F_\varepsilon^{-1}(\tau), \hat{\beta}_2, \ldots, \hat{\beta}_p)^T.$$

There is no need for quantile regression under these models, because the conditional mean and some associated measure of dispersion have already better properties. In real data analysis this case is rare. We observe rather errors, which are long tailed or the model is heteroscedastic or a mixture of both. Then either a robust alternative to the least squares approach or a heteroscedastic extension of the model (2.3) is needed. For the first case the conditional median is a good and already well known alternative to the conditional mean which lacks the robustness in case of heavy tailed or even only skewed error distributions

$$\widehat{median}(Y|x) = x^T \hat{\beta}_{0.5}.$$

In the case of heteroscedastic error terms an extension to the linear model in (2.3) is the linear location scale-model

$$Y_i = x_i^T \beta + (x_i^T \gamma)\varepsilon_i, \quad i = 1, 2, \ldots, n \tag{2.9}$$

with $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)^T$ being another unknown p-dimensional parameter vector. In this model the conditional quantile is given as

$$Q_{Y_i|x_i}(\tau) = x_i^T \beta + x_i^T \gamma \cdot F_\varepsilon^{-1}(\tau), \quad i = 1, 2, \dots, n,$$

leading to representation of $\hat{\beta}_\tau$ depending on the mean estimator $\hat{\beta}$ in (2.4)

$$\hat{\beta}_\tau = (\hat{\beta}_1 + \hat{\gamma}_1 \cdot F_\varepsilon^{-1}(\tau), \hat{\beta}_2 + \hat{\gamma}_2 \cdot F_\varepsilon^{-1}(\tau), \dots, \hat{\beta}_p + \hat{\gamma}_p \cdot F_\varepsilon^{-1}(\tau))^T.$$

This implies that all parameters $x_{i,p}$ have the same monotone behaviour in τ governed by $F_\varepsilon^{-1}(\tau)$. Clearly this is too restrictive. On top the distribution of ε_i, $i = 1, 2, \dots, n$, is often unknown. In order to be more flexible in terms of the distribution of the error terms and dependencies of them with the several parameters in x Koenker and Bassett [1978] introduced the linear quantile model

$$Q_{Y_i|x_i}(\tau) = x_i^T \beta_\tau, \quad i = 1, 2, \dots, n. \tag{2.10}$$

Hence the conditional quantile is expressed as a linear combination of the auxiliary variable x. In this model there is no implied interconnection between β from the linear model (2.3) and β_τ.

2.3.1 *Expectiles and M-quantiles*

As mentioned before the median is a robust alternative to the mean. There are other robust estimators which can be described by the employed loss function. Similar to quantiles the estimation of *expectiles* uses a quadratic loss function. Newey and Powell [1987] introduced this idea. There the 0.5-*expectile* is the mean. For other τ the estimator is not easily interpreted. Breckling and Chambers [1988] introduced *M-quantiles* which represent estimators between *expectiles* and quantiles. There the loss function is a variable Huber function (Huber proposal II) with a parameter c which can be chosen. Thus based on this choice *expectiles* and quantiles are special cases of *M-quantiles*. Therefore the 0.5-*M-quantile* is another robust alternative to the mean. Chambers and Tzavidis [2006] firstly used *M-quantiles* in *Small Area Estimation*.

2.4 THE EQUIVALENCE TO ASYMMETRIC LAPLACIAN MODEL

In the following I am going to show that the minimising problem as introduced in (2.6) is equivalent to a *maximum likelihood* estimation. The likelihood function of this model comes from an *asymmetric Laplace distribution*, which I will discuss before in the following Section.

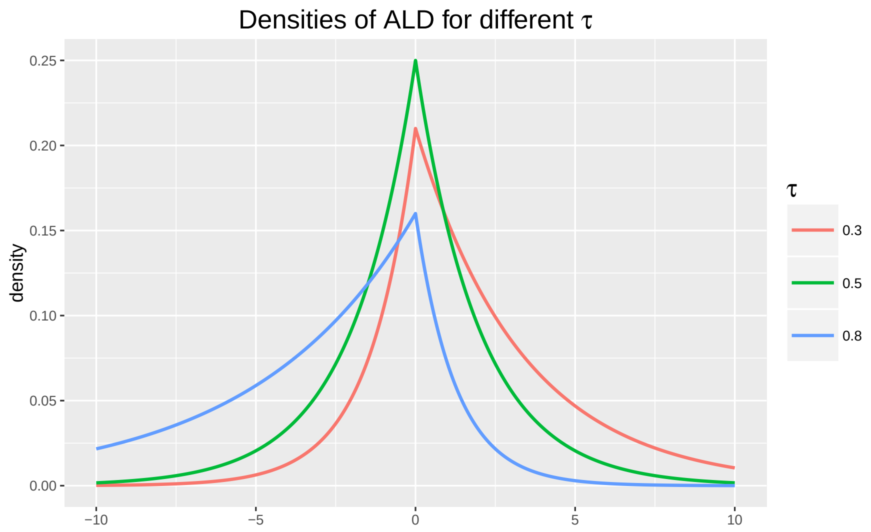

Figure 2.1: Densities of the *asymmetric Laplace distribution* with $\mu = 0$, $\sigma = 1$, and $\tau = 0.3, 0.5, 0.8$.

2.4.1 *The Asymmetric Laplace Distribution*

If a random variable Y is distributed according to the *asymmetric Laplace distribution*, $Y \sim \text{ALD}(\mu, \sigma, \tau)$, its density function is given by

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right) \qquad (2.11)$$

with $\rho_\tau(\cdot)$ as introduced in (2.2). One of the earliest appearances in literature on the *asymmetric Laplace distribution* can be found in McGill [1962]. It is characterised by three parameters μ, σ, and τ. $\mu \in \mathbb{R}$ is the location parameter. If μ was increased by a number a the density is shifted on the x-axis by this value. $\sigma \in \mathbb{R}_+$ is the scale parameter. For higher σ the density is wider and the data is more spread. $\tau \in (0, 1)$ is the skewness parameter. For $\tau = 0.5$ the density is symmetric around μ and it is the *double Laplace distribution* which is a *Laplace distribution* with location parameter μ and scale parameter 2σ (cf. Laplace [1774]). For $\tau < 0.5$ the distribution is positively skewed and for $\tau > 0.5$ it is negatively skewed. Figure 2.1 shows the density for $\mu = 0$, $\sigma = 1$, and different values of τ. There it can be seen that the point of non-differentiability is at $\mu = 0$ and the distribution is positively skewed for $\tau = 0.3$ and negatively skewed for $\tau = 0.8$.

Yu and Zhang [2005] discussed the distribution and its properties in detail. A basic property of it is that

$$P(Y \leqslant \mu) = \tau$$

and thus that the position parameter μ is the τ-quantile of the distribution

$$Q_Y(\tau) = \mu.$$

As a result for a given $\tau \in (0, 1)$ any estimator for μ is a τ-quantile estimator

$$\hat{\mu} = \hat{Q}_Y(\tau).$$

Furthermore Yu and Zhang [2005] showed the following property

Lemma 2.1. *The k^{th} central moment of an asymmetric Laplacian distributed $Y \sim ALD(\mu, \sigma, \tau)$ is given by*

$$E\left[(Y-\mu)^k\right] = k!\sigma^k \tau(1-\tau)\left(\frac{1}{\tau^{k+1}} + \frac{(-1)^k}{(1-\tau)^{k+1}}\right). \qquad (2.12)$$

From these moments I can derive other expectations and variances of an *asymmetric Laplacian distributed* random variable Y. Therefore let me state the following corollary

Corollary 2.2. *For a random variable $Y \sim ALD(\mu, \sigma, \tau)$ there holds the following.*

(a) The expected value is given as

$$E[Y] = \mu + \frac{\sigma(1-2\tau)}{\tau(1-\tau)}.$$

(b) The variance is given as

$$Var(Y) = \frac{\sigma^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2}.$$

(c) The expected value of $Y\mathbf{1}_{\{Y \leqslant \mu\}}$ is given as

$$E\left[Y\mathbf{1}_{\{Y \leqslant \mu\}}\right] = \tau\mu - \frac{\tau\sigma}{1-\tau}.$$

(d) The expected value of $Y^2\mathbf{1}_{\{Y \leqslant \mu\}}$ is given as

$$E\left[Y^2\mathbf{1}_{\{Y \leqslant \mu\}}\right] = \tau\mu^2 - 2\frac{\tau\sigma\mu}{1-\tau} + \frac{2\tau\sigma^2}{(1-\tau)^2}.$$

(e) The expected value of $\mathbf{1}_{\{Y \leqslant \mu\}}$ is given as

$$E\left[\mathbf{1}_{\{Y \leqslant \mu\}}\right] = \tau.$$

Before I am going to prove the properties let me remark that the expectation and variance of the *asymmetric Laplace distribution* in (a) and (b) are already well known. Nevertheless I am calculating them from the lemma stated before as an exercise. The other rather odd looking statements are later needed in the main proof of this thesis. Since I employ properties of the *asymmetric Laplace distribution* I show them at this point of the thesis.

Proof. (a) By using (2.12) for $k = 1$ the expected value can be calculated as follows

$$\begin{aligned}
E[Y] &= E\left[(Y-\mu)^1\right] + \mu \\
&= \sigma\tau(1-\tau)\left(\frac{1}{\tau^2} + \frac{-1}{(1-\tau)^2}\right) + \mu \\
&= \mu + \sigma\tau(1-\tau)\frac{(1-\tau)^2 - \tau^2}{\tau^2(1-\tau)^2} \\
&= \mu + \sigma\frac{(1-2\tau+\tau^2) - \tau^2}{\tau(1-\tau)} \\
&= \mu + \frac{\sigma(1-2\tau)}{\tau(1-\tau)}.
\end{aligned}$$

(b) By using (2.12) for $k = 2$ the variance can be calculated as follows

$$\begin{aligned}
Var(Y) &= Var(Y-\mu) \\
&= E\left[(Y-\mu)^2\right] - E^2[Y-\mu] \\
&= 2\sigma^2\tau(1-\tau)\left(\frac{1}{\tau^3} + \frac{1}{(1-\tau)^3}\right) \\
&\quad - \left(\sigma\tau(1-\tau)\left(\frac{1}{\tau^2} + \frac{-1}{(1-\tau)^2}\right)\right)^2 \\
&= 2\sigma^2\frac{(1-\tau)^3 + \tau^3}{\tau^2(1-\tau)^2} \\
&\quad - \left(\sigma\frac{(1-\tau)^2 - \tau^2}{\tau(1-\tau)}\right)^2 \\
&= \frac{2\sigma^2(1-3\tau+3\tau^2) - \sigma^2(1-2\tau)^2}{\tau^2(1-\tau)^2} \\
&= \frac{\sigma^2(2-6\tau+6\tau^2) - \sigma^2(1-4\tau+4\tau^2)}{\tau^2(1-\tau)^2} \\
&= \frac{\sigma^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}.
\end{aligned}$$

(c) With applications of *integration by parts* $(\star)$ the expression can be calculated as follows

$$\begin{aligned}
E\left[Y\mathbf{1}_{\{Y\leqslant\mu\}}\right] &= \int_{-\infty}^{\mu} y f(y)\,dy \\
&= \int_{-\infty}^{\mu} y\frac{\tau(1-\tau)}{\sigma}\exp\left(-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right)dy \\
&= \int_{-\infty}^{\mu} y\frac{\tau(1-\tau)}{\sigma}\exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)dy
\end{aligned}$$

$$
\begin{aligned}
&\overset{(\star)}{=} \tau y \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\Big|_{-\infty}^{\mu} \\
&\quad - \int_{-\infty}^{\mu} \tau \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&= \tau\mu \exp(0) - \tau \lim_{y\to-\infty} y \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) \\
&\quad - \frac{\tau\sigma}{1-\tau} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\Big|_{-\infty}^{\mu} \\
&\overset{(\star\star)}{=} \tau\mu - 0 \\
&\quad - \frac{\tau\sigma}{1-\tau}\left(\exp(0) - \lim_{y\to-\infty} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\right) \\
&= \tau\mu - \frac{\tau\sigma}{1-\tau},
\end{aligned}
$$

where $(\star\star)$ is an application of *L'Hôpital's rule*.

(d) With twice the application of *integration by parts* $(\star)$ the expression can be calculated a s follows

$$
\begin{aligned}
E\left[Y^2 \mathbf{1}_{\{Y\leqslant 0\}}\right] &= \int_{-\infty}^{\mu} y^2 f(y) dy \\
&= \int_{-\infty}^{\mu} y^2 \frac{\tau(1-\tau)}{\sigma} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&\overset{(\star)}{=} \tau y^2 \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\Big|_{-\infty}^{\mu} \\
&\quad - \int_{-\infty}^{\mu} 2\tau y \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&\overset{(\star\star)}{=} \tau\mu^2 - 2\int_{-\infty}^{\mu} \tau y \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&\overset{(\star)}{=} \tau\mu^2 - 2\frac{\tau\sigma}{1-\tau} y \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\Big|_{-\infty}^{\mu} \\
&\quad + 2\int_{-\infty}^{\mu} \frac{\tau\sigma}{1-\tau} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&\overset{(\star\star)}{=} \tau\mu^2 - 2\frac{\tau\sigma\mu}{1-\tau} \\
&\quad + \frac{2\tau\sigma^2}{(1-\tau)^2}\left(\exp(0) - \lim_{y\to-\infty} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\right) \\
&= \tau\mu^2 - 2\frac{\tau\sigma\mu}{1-\tau} + \frac{2\tau\sigma^2}{(1-\tau)^2},
\end{aligned}
$$

where $(\star\star)$ is an application of *L'Hôpital's rule* which was applied twice in the first time.

(e) The expected value of $\mathbb{1}_{\{Y \leqslant \mu\}}$ can be rewritten as a probability which is calculated as follows

$$\begin{aligned}
E\left[\mathbb{1}_{\{Y \leqslant \mu\}}\right] &= P(Y \leqslant \mu) \\
&= \int_{-\infty}^{\mu} f(y)dy \\
&= \int_{-\infty}^{\mu} \frac{\tau(1-\tau)}{\sigma} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right) dy \\
&= \tau \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\Bigg|_{-\infty}^{\mu} \\
&= \tau\left(1 - \lim_{y \to -\infty} \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right)\right) \\
&= \tau.
\end{aligned}$$

□

As mentioned before some of the expressions in Corollary 2.2 are derived because they are needed later in the consistency proof in Chapter 4. Now let us examine the standard *asymmetric Laplace distribution*. A random variable X is said to be *standard asymmetric Laplacian distributed* if

$$X \sim ALD(0, 1, \tau).$$

Any *asymmetric Laplacian distributed* random variable Y can be derived from the *standard asymmetric Laplacian distributed* random variable X through the transformation

$$Y = \mu + \sigma X \sim ALD(\mu, \sigma, \tau). \tag{2.13}$$

For a vector of independently *asymmetric Laplacian distributed* random variables $Y = (Y_1, Y_2, \dots, Y_n)^T$ with $Y_i \overset{i}{\sim} ALD(\mu_i, \sigma, \tau)$ the density is given by

$$f(y) = \frac{\tau^n(1-\tau)^n}{\sigma^n} \exp\left(\sum_{i=1}^{n}\left(-\rho_\tau\left(\frac{y_i-\mu_i}{\sigma}\right)\right)\right). \tag{2.14}$$

There the scale parameter σ and the skewness parameter τ are assumed to be the same for all Y_i, $i = 1, 2, \dots, n$. The location parameters μ_i may be different though. We write for Y

$$Y \sim ALD_n(\boldsymbol{\mu}, \sigma, \tau)$$

with $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$.

In linear modelling the *asymmetric Laplace distribution* can be used as the error distribution leading to the *asymmetric Laplacian linear model* given by

$$Y_i = x_i^T \beta_\tau + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{2.15}$$

with

$$\varepsilon_i \overset{iid}{\sim} \text{ALD}(0, \sigma, \tau).$$

This model leads by (2.13) to the conditional distribution of Y_i given x_i

$$Y_i|x_i \overset{iid}{\sim} \text{ALD}(x_i^T \beta_\tau, \sigma, \tau).$$

The *asymmetric Laplacian linear model* (2.15) is mainly applied in quantile estimation (cf. e.g. Koenker [2005] or Machado and Santos Silva [2005]), because it is equivalent to the linear quantile model (2.10) which is shown in the following Section.

2.4.2 *The Equivalence of the Models*

In the following I am going to show the equivalence of the linear quantile model (2.10) to the linear model with asymmetric Laplacian error terms (2.15). This is important for the estimation process on the one hand and for the asymptotic performance of the estimator on the other hand. The minimising problem from (2.6) is shifted to a *maximum likelihood* estimation in the *asymmetric Laplacian* model. Either way the estimation of β_τ can be fulfilled and its consistency shown, which will be discussed in Section 2.5. The field of *maximum likelihood* estimation is well investigated. For this reason the equivalence of the two models is proven in Theorem 2.3. The equivalence has already been employed in many quantile regression applications (cf. e.g. Machado and Santos Silva [2005]). However in order to give the reader a deeper understanding of the theory behind I will conduct this proof here.

For random variables $Y_i|x_i \overset{iid}{\sim} \text{ALD}(\mu_i, \sigma, \tau)$ with $\mu_i = x_i^T \beta_\tau$ the likelihood density function is given for $\mu = (\mu_1, \mu_2 \ldots, \mu_n)^T$ as a density of an n-dimensional *asymmetric Laplace distribution*

$$\begin{aligned} L(\boldsymbol{\mu}, \sigma, \tau, Y) &= \prod_{i=1}^{n} f_{\text{ALD}(\mu_i, \sigma, \tau)} \\ &\overset{(\star)}{=} \prod_{i=1}^{n} \frac{\tau(1-\tau)}{\sigma} \exp\left(-\rho_\tau\left(\frac{Y_i - \mu_i}{\sigma}\right)\right) \\ &= \frac{\tau^n (1-\tau)^n}{\sigma^n} \exp\left(-\sum_{i=1}^{n} \rho_\tau\left(\frac{Y_i - \mu_i}{\sigma}\right)\right) \end{aligned}$$

where $(\star)$ follows from (2.14). For a fixed $\tau \in (0, 1)$ it is proportional to

$$L(\boldsymbol{\mu}, \sigma, \tau, Y) \propto \sigma^{-n} \exp\left(-\sum_{i=1}^{n} \rho_\tau\left(\frac{Y_i - \mu_i}{\sigma}\right)\right).$$

The *maximum likelihood* estimator for $\mu_i = x_i^T \beta_\tau$ is given by

$$\hat{\mu}_i^{MLE} = x_i^T \hat{\beta}_\tau^{MLE}$$

with

$$\hat{\beta}_\tau^{MLE}(Y|\mathbf{X}) := \underset{\beta_\tau \in \mathbb{R}^p}{\arg\max} \left\{ \sigma^{-n} \exp\left(-\sum_{i=1}^{n} \rho_\tau \left(\frac{Y_i - x_i^T \beta_\tau}{\sigma} \right) \right) \right\}. \tag{2.16}$$

In the research field Jung [1996] introduced already a quasi-maximum likelihood approach in median regression using a *Laplace distribution*. Then Koenker and Machado [1999] mentioned the *asymmetric Laplace distribution* for quantile regression in a frequantistic approach. Yu and Moyeed [2001] mentioned the use of the of the *asymmetric Laplace distribution* in a Bayesian approach. In the following theorem I will prove the equivalence of the quantile model as introduced in (2.10) to the asymmetric Laplacian model in (2.15) by showing the equality of their estimators for the parameter vector β_τ.

Theorem 2.3. *For a fixed* $\tau \in (0, 1)$ *the estimators* $\hat{\beta}_\tau(Y|\mathbf{X})$ *from (2.7) and* $\hat{\beta}_\tau^{MLE}(Y|\mathbf{X})$ *from (2.16) coincide*

$$\hat{\beta}_\tau(Y|\mathbf{X}) = \hat{\beta}_\tau^{MLE}(Y|\mathbf{X}).$$

Proof. By (2.7) and (2.16) it holds that

$$\begin{aligned}
\hat{\beta}_\tau(Y|\mathbf{X}) &= \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau(Y_i - x_i^T \beta_\tau) \\
&= \underset{\beta \in \mathbb{R}^p}{\arg\max} \left\{ -\sum_{i=1}^{N} \rho_\tau(Y_i - x_i^T \beta_\tau) \right\} \\
&= \underset{\beta \in \mathbb{R}^p}{\arg\max} \left\{ \sigma^{-n} \exp\left(-\sum_{i=1}^{n} \rho_\tau \left(\frac{Y_i - x_i^T \beta_\tau}{\sigma} \right) \right) \right\} \\
&= \hat{\beta}_\tau^{MLE}(Y|\mathbf{X}).
\end{aligned}$$

□

This proven equivalence of the quantile linear model (2.10) to the *asymmetric Laplacian linear model* (2.15) transfers the minimising problem from (2.6) to the problem field of *maximum likelihood* estimation. The equivalent model to (2.10) is now model (2.15)

$$Q_{Y_i|x_i}(\tau) = x_i^T \beta_\tau \quad \Leftrightarrow \quad Y_i = x_i^T \beta_\tau + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{2.17}$$

with

$$\varepsilon_i \overset{iid}{\sim} ALD(0, \sigma, \tau), \quad i = 1, 2, \dots, n.$$

Hence the estimation of β_τ can be derived by both methods as described in (2.7) and (2.16). Besides asymptotic theory can be studied on both approaches as I will discuss in Section 2.5. Note that this equivalence is only numerical. The *asymmetric Laplace distribution* for the error term is an approach in order to transfer the estimation by minimising the risk function into a *maximum likelihood* problem. Latter is quite well explored. Actually the *asymmetric Laplace distribution* was only constructed for quantile estimation. Let me go into detail of the linear quantile model with *asymmetric Laplacian distributed* error terms by discussing the meaning of the unknown scale parameter σ before.

2.4.3 *The Meaning of* σ

The scale parameter σ of the *asymmetric Laplace distribution* has to be positive ($\sigma > 0$). If we set it to a fixed number, the equivalence of the models in Theorem 2.3 would still hold. Instead of setting it fixed though, I keep σ, the scale parameter of the error term ε, as an unknown parameter. This gives more flexibility in the estimation whenever the model does not fit the reality. In other publications, for example in Koenker and Machado [1999], σ is fix and set equal to one. So by extending the *asymmetric Laplacian model* with this change the equivalence in (2.17) still holds and we have more flexibility in terms of variation in the data. The new model can then be written as

$$Y_i = x_i^T \beta_\tau + \varepsilon_i, \quad i = 1, 2, \dots, n, \tag{2.18}$$

where

$$\varepsilon_i \overset{iid}{\sim} \mathrm{ALD}(0, \sigma, \tau) \quad \text{with } \sigma > 0 \text{ unknown.}$$

Since we assume that σ is unknown, it can be estimated simultaneously to β_τ leading to the estimators

$$\begin{aligned}(\hat{\beta}_\tau^T(Y|\mathbf{X}), \hat{\sigma}(Y|\mathbf{X}))^T \\ := \underset{(\beta^T,\sigma)^T \in \mathbb{R}^p \times \mathbb{R}_+}{\arg\max} \left\{ \sigma^{-n} \exp\left(-\sum_{i=1}^{n} \rho_\tau \left(\frac{Y_i - x_i^T \beta_\tau}{\sigma} \right) \right) \right\}.\end{aligned} \tag{2.19}$$

The estimation of σ may be useful in testing problems related to the regression because it determines the distribution of the test statistic.

2.5 THE CONSISTENCY OF QUANTILE REGRESSION IN LINEAR MODELS

For the terminology consistency let me state the following definition according to the definition from Georgii [2009].

Definition 2.4. *Let* Y *be an* n*-dimensional vector of observations. An estimator* $\hat{S}(Y)$ *is consistent for the true value* S^0 *if and only if for all* $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\hat{S}(Y) - S^0\right| \geqslant \epsilon\right) = 0.$$

The consistency of quantile regression in the linear model has been proven in several different ways. Many of them not only show the consistency but the *asymptotic normality* (cf. Chapter 9.3 of van der Vaart [2007]) of the parameter estimator $\hat{\beta}_\tau$. For independently distributed observations $Y_1, Y_2, \ldots, Y_n$ the classical *maximum likelihood* theory is applicable. There the *Fisher Information* matrix plays an important role.

Definition 2.5. *For independently distributed observations* $Y_1, Y_2, \ldots, Y_n$ *each with density* $f(y|\theta)$ *the Fisher Information matrix at* θ^0 *is defined as*

$$I(\theta^0) := E\left[\frac{\partial}{\partial\theta}\log\left(f(y_i|\theta)\right)\frac{\partial}{\partial\theta}\log\left(f(y_i|\theta)\right)^T\Big|_{\theta^0}\right]. \tag{2.20}$$

Let me state the following theorem for *maximum likelihood* estimators of *Glivenko-Cantelli* kind (cf. Glivenko [1933] and Cantelli [1933]).

Theorem 2.6. *Let* $Y_1, Y_2, \ldots, Y_n$ *be iid, each with density* $f(y|\theta)$ *where* $\theta \in \Theta$ *and suppose that the following regularity assumptions hold.*

(a) The parameter space Θ *is compact and the real value* θ^0 *is in the interior of* Θ*.*

(b) For the distributions holds $P_\theta \neq P_{\theta^0}$ *for all* $\theta \neq \theta^0$*.*

(c) For every y *the density* $f(y|\theta)$ *is twice differentiable with respect to* θ *and the second derivative is continuous in an environment of* θ^0*.*

(d) There are dominating random variable M_0, M_1, M_2 *such that* $\sup_\theta \log(f(y|\theta)) \leqslant M_0(y)$*,* $\sup_\theta \frac{\partial}{\partial\theta}\log(f(y|\theta)) \leqslant M_1(y)$*, and* $\sup_\theta \frac{\partial^2}{\partial\theta^2}\log(f(y|\theta)) \leqslant M_2(y)$*.*

(e) The Fisher Information $I(\theta)$ *defined by Definition 2.5 is positive definite.*

Then the maximum likelihood estimator $\hat{\theta}_n$ *is under* $P^n_{\theta^0}$ *asymptotically normal*

$$\sqrt{n}\left(\hat{\theta}_n - \theta^0\right) \xrightarrow{\mathcal{D}} N\left(0, I^{-1}(\theta^0)\right).$$

Proof. A proof can be found in Lehmann and Casella [1998] Theorem 3.10. □

Huber [1967] studied the *maximum likelihood* estimation in non-standard conditions. There he examined the performance whenever the regularity assumptions are violated. The latter theorem can be directly applied on quantile estimation. Let $Y_1, Y_2, \ldots, Y_n$ be iid observations with continuous density f and distribution function F and

$\tau \in (0,1)$ is fixed. As discussed before the τ-quantile of Y can be estimated by finding a *maximum likelihood* estimator for μ. It follows by application of Theorem 2.6 that

$$\sqrt{n}(\hat{Q}_Y(\tau) - Q_Y(\tau)) \xrightarrow{\mathcal{D}} N\left(0, \left(\frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right)^2\right).$$

For details see Koenker [2005] Chapters 3.2 and 4.1. Thus the rate of convergence is the classical $\sqrt{n}$-rate. Note that $\tau(1-\tau)$ has its minimum for $\tau = 0.5$, which leads to the lowest variance for the median estimation. Hence the estimation of the median outperforms the estimation of other quantiles in the sense that its asymptotic variance is the smallest. This is intuitive because quantiles in the lower or higher regions are more prone to the few observations in the edges.
For quantile regression in the linear model the approach is similar and Koenker [2005] states the following theorem.

Theorem 2.7. *Let $\tau \in (0,1)$ be fixed and let us assume the linear quantile model 2.10 with conditional distribution functions $F_{Y_i|x_i}$ and the estimator $\hat{\beta}_\tau$ from (2.7). Suppose that the following conditions hold.*

(a) The distribution functions $F_{Y_i|x_i}$ are absolutely continuous with continuous densities $f_{Y_i|x_i}$ uniformly bounded away from zero and infinity at the points $Q_{Y_1|x_1}(\tau), Q_{Y_2|x_2}(\tau), \ldots, Q_{Y_n|x_n}(\tau)$.

(b) There exists a positive definite matrix D_0 such that $\frac{1}{n}\lim_{n\to\infty}\sum_{i=1}^n x_i x_i^T = D_0$.

(c) There exists a positive definite matrix D_1 such that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_{Y_i|x_i}(Q_{Y_i|x_i}(\tau))x_i x_i^T = D_1$.

(d) $\max_{i=1,2,\ldots,n}\frac{\|x_i\|}{\sqrt{n}} \to 0$.

Then as $n \to \infty$

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{\mathcal{D}} N\left(0, \tau(1-\tau)D_1^{-1}D_0D_1^{-1}\right).$$

Proof. See Theorem 4.1. in Koenker [2005]. □

In the proof of the latter theorem the equivalence to a *maximum likelihood* estimation was not used. It was directly proved for $\hat{\beta}_\tau$ from (2.7). However by applying Theorem 1 from Pollard [1991] on the linear quantile model 2.10 I get the same result. There the proof is based on the *maximum likelihood* approach. Nevertheless, *asymptotic normality* implies consistency which can be shown by the following small proof:
Let the estimator $\hat{\theta}$ be asymptotically normal with rate $\sqrt{n}$ and asymptotic covariance matrix A. θ^0 is the real value. Thus it holds

$$\sqrt{n}(\hat{\theta} - \theta^0) \xrightarrow{\mathcal{D}} N(0, A).$$

It holds for an arbitray $\epsilon > 0$

$$P\left(\left|\hat{\theta} - \theta^0\right| \geqslant \epsilon\right) = P\left(\sqrt{n}A^{-\frac{1}{2}}\left|\hat{\theta} - \theta^0\right| \geqslant \sqrt{n}A^{-\frac{1}{2}}\epsilon\right).$$

Due to the asymptotic normality the asymptotic distribution of $\sqrt{n}A^{-\frac{1}{2}}\left|\hat{\theta} - \theta^0\right|$ is a standard normal whilst the term on the right hand side $\sqrt{n}A^{-\frac{1}{2}}\epsilon$ diverges to infinity as $n \to \infty$. Thus the probability converges to zero which implies the consistency.
This implication leads to the following corollary.

Corollary 2.8. *The estimator $\hat{\beta}_\tau$ – see (2.7) or (2.16) – in the linear quantile model (2.10) is consistent.*

Proof. Set $A := \tau(1-\tau)D_1^{-1}D_0D_1^{-1}$. Let $\epsilon > 0$ be arbitrary and fixed

$$\begin{aligned}
&P\left(\left|\hat{\beta}_\tau - \beta_\tau\right| \geqslant \epsilon\right) \\
&\quad = P\left(\sqrt{n}A^{-\frac{1}{2}}\left|\hat{\beta}_\tau - \beta_\tau\right| \geqslant \sqrt{n}A^{-\frac{1}{2}}\epsilon\right) \\
&\quad \overset{(\star)}{\to} 2\left(1 - \phi\left(\sqrt{n}A^{-\frac{1}{2}}\epsilon\right)\right) \\
&\quad \overset{(\star\star)}{\to} 0 \quad \text{as } n \to \infty,
\end{aligned}$$

where $(\star)$ follows because $\sqrt{n}A^{-\frac{1}{2}}(\hat{\beta}_\tau - \beta_\tau)$ is by Theorem 2.7 *asymptotically standard normal* distributed and ϕ is its probability density function. Due to the definition of $\sqrt{n} \to \infty$ $(\star\star)$ follows because $\phi(\sqrt{n}\cdot C) \to 1$ as $n \to \infty$ for any constant C. As a remark, the convergence $(\star)$ is just a between step of the whole convergence proven. This is why at this stage the limit still depends on n. □

The consistency of $\hat{\beta}_\tau$ implies the consistency of the whole quantile estimator $\hat{Q}_{Y_i|x_i}(\tau)$ in (2.8).
In further research Bassett and Koenker [1986] proved a strong consistency of quantile regression in linear models and Buchinsky [1995] showed a method of computing the covariance matrix via a *Monte Carlo* approach. Powell [1983] showed the *asymptotic normality* for two-stage least absolute deviation estimators such as the median.

2.6 FURTHER RESEARCH ON QUANTILE REGRESSION IN LINEAR MODELS

After Koenker and Bassett [1978] introduced the quantile regression in linear models, further research based on this theory was made. For example can be observed that quantile regression lines for different $\tau \in (0, 1)$ may cross. This happens especially in the edges of the support of the auxiliary variables x because of fewer observations in this part. Of course this is theoretically impossible under the assumption of the model. However He [1997] introduced a method to solve the

crossing quantiles problem.
Rogers [2001] discussed least absolute deviation estimators in non-standard conditions. There he focusses mainly on applications in time series. Koenker [2004] already introduced quantile regression for longitudinal data using a fixed effects model.
Jones [1994] showed that there is a link between expectiles and quantiles. Expectiles are estimators with the *Euclidean distance* as loss function with a skewness parameter $\tau \in (0,1)$. The 0.5-expectile is the mean. Thus he instructs a way of computing medians from means and vice versa.

2.7 APPLICATIONS AND EXTENSIONS OF THE LINEAR QUANTILE MODEL

Naturally the linear quantile model can be applied, whenever we are interested in quantiles of dependent variables $Y_i, Y_2, \dots, Y_n$, which follow the linear model (2.3) with unknown error distribution. In practice this model is often only a starting point and can be improved. In mean estimation this is already well investigated. There the generalised linear models were introduced employing a link function on the data in order to get back to a linear modelling problem. In Section 2.7.1 I will show that this also works with quantiles leading to a generalised linear quantile model. A short guidance to the estimation is given and the property of consistency can even be stated for the quantile estimators in these cases. Another extension of the quantile regression in linear models was discussed by Machado and Santos Silva [2005]. They introduced a method for estimating quantiles of counts and showed the consistency of their introduced estimator. This I will present in Section 2.7.2.

2.7.1 *Generalised Linear Quantile Models*

For data $Y_i, Y_2, \dots, Y_n$ following a linear model (2.3) with unknown error distribution the linear quantile modelling is directly applicable. The parameter β_τ can be estimated either by minimising the lost function (2.6) or as a *maximum likelihood* estimator in the linear quantile model (2.18) with *asymmetric Laplacian* error terms. The results are the same as shown in Theorem 2.3 leading to conditional quantile estimators for a given $\tau \in (0,1)$

$$Q_{Y_i|x_i}(\tau) = x_i^T \hat{\beta}_\tau \quad i = 1, 2, \dots, n.$$

In cases of a generalised linear model a step in the estimating process has to be added. Let us assume that the mean model is given as

$$T\left(E\left[Y_i\right]\right) = x_i^T \beta, \quad i = 1, 2, \dots, n,$$

or equivalently

$$E[Y_i] = T^{-1}\left(x_i^T \beta\right), \quad i = 1, 2, \ldots, n. \tag{2.21}$$

Here the link function $T(\cdot)$ is a continuous function. An example may be the log-linear model

$$\log\left(E[Y_i]\right) = x_i^T \beta, \quad i = 1, 2, \ldots, n,$$

or equivalently

$$E[Y_i] = \exp\left(x_i^T \beta\right), \quad i = 1, 2, \ldots, n.$$

With the same argumentation as in the linear model in Section 2.3 the model (2.21) may be used for quantile estimation. For a fixed $\tau \in (0, 1)$ this leads to the generalised linear quantile model

$$Q_{Y_i|x_i}(\tau) = T^{-1}\left(x_i^T \beta_\tau\right), \quad i = 1, 2, \ldots, n, \tag{2.22}$$

or equivalently

$$T\left(Q_{Y_i|x_i}(\tau)\right) = x_i^T \beta_\tau, \quad i = 1, 2, \ldots, n.$$

For a continuous transformation function T it holds that the transformation of the quantile is the quantile of the transformation

$$T\left(Q_{Y_i|x_i}(\tau)\right) = Q_{T(Y_i)|x_i}(\tau), \quad i = 1, 2, \ldots, n. \tag{2.23}$$

In application for observations $Y_1, Y_2, \ldots, Y_n$ following the generalised linear model (2.21) we need to transform the data in a first step.

$$Y_i \to T(Y_i), \quad i = 1, 2, \ldots, n.$$

On the transformed observations $T(Y_1), T(Y_2), \ldots, T(Y_n)$ the linear quantile model (2.10) holds. In the second step we can estimate $\hat{\beta}_\tau$ leading to conditional quantile estimators of them

$$\hat{Q}_{T(Y_i)|x_i}(\tau) = x_i^T \hat{\beta}_\tau, \quad i = 1, 2, \ldots, n.$$

This estimation is by Corollary 2.8 consistent. In the third step the transformed conditional quantile estimators of the observations $Y_1, Y_2, \ldots, Y_n$ are by (2.23) given as

$$T\left(\hat{Q}_{Y_i|x_i}(\tau)\right) = \hat{Q}_{T(Y_i)|x_i}(\tau) = x_i^T \hat{\beta}_\tau, \quad i = 1, 2, \ldots, n,$$

which is eventually equivalent to their quantile estimator

$$\hat{Q}_{Y_i|x_i}(\tau) = T^{-1}\left(x_i^T \hat{\beta}_\tau\right), \quad i = 1, 2, \ldots, n.$$

Since T is continuous, the estimation is still consistent.

2.7.2 *Linear Quantile Models for Count Data*

In mean estimation a Poisson model is often used with count data. However, this model is prone to data which suffers an overdispersion. Efron [1992] solved the problem by using an asymmetric *maximum likelihood* approach. Another way could be the estimation of parameter which are less prone to sparse data, such as the median estimator. Other quantiles may also be of interest. The estimation of quantiles for count data can not be straightforwardly derived by applying the linear quantile model. Quantiles of count data must be integers due to the fact that counts themselves are integers. Since the linear quantile model (2.10) is a model for continuous data, it is not directly applicable on counts. The count mean model or *Poisson* model for a discrete random variable is Y_i given x_i can be stated as

$$\exp(x_i^T \beta), \quad i = 1, 2, \dots, n. \tag{2.24}$$

This mean model needs to be improved in order to estimate quantiles of Y_i given x_i for a fixed $\tau \in (0, 1)$, $Q_{Y_i|x_i}(\tau)$. The main idea in estimating quantiles for counts was developed by Machado and Santos Silva [2005]. They applied a method called *jittering* on the observed random variables $Y_1, Y_2, \dots, Y_n$, which will be discussed in the following Section 2.7.2.1. Later I state the main theorem of their article in Section 2.7.2.5. There I state the consistency of the count quantiles derived through the *jittering* approach.

2.7.2.1 *Jittering the Count Data*

The count observations Y_i ($i = 1, 2, \dots, n$) are discrete. Machado and Santos Silva [2005] introduced the idea of *jittering* in order to get continuous data. This method means adding a *standard uniform* random variable U_i independent from Y_i and x_i, we get a continuous observation Z_i:

$$Z_i := Y_i + U_i. \tag{2.25}$$

On this continuous random variable Z_i we can apply the linear quantile model (2.10). Its quantile can be stated in the following theorem.

Theorem 2.9. *For a fixed $\tau \in (0, 1)$ the quantile of Z_i as defined in (2.25) is said to be*

$$Q_{Z_i|x_i}(\tau) = \exp(x_i^T \beta) + \tau.$$

Proof. Let $\tau \in (0, 1)$ be fixed. For a continuous random variable $Y_i + U(-\tau, 1-\tau)$, where the mean model (2.24) holds for Y_i the τ-quantile is

$$\begin{aligned} & Q_{Y_i+U(-\tau,1-\tau)|x_i}(\tau) = \exp(x_i^T\beta) \\ \iff \quad & Q_{Y_i+U(-\tau,1-\tau)+\tau|x_i}(\tau) = \exp(x_i^T\beta) + \tau \\ \iff \quad & Q_{Y_i+U(0,1)|x_{ij}}(\tau) = \exp(x_i^T\beta) + \tau. \end{aligned}$$

□

2.7.2.2 *Transformation of the Jittered Data*

In order to be able to apply the quantile estimation approach of linear quantile models (2.10), there is need to transform the jittered data Z_i. This is for a fixed $\tau \in (0, 1)$ fulfilled as follows

$$T(Z_i, \tau) := \begin{cases} \log(\zeta), Z_i \leqslant \tau \\ \log(Z_i - \tau), Z_i > \tau \end{cases}$$

with a small value ζ. This transformation is almost a continuous function and $\log(\zeta)$ is just the function value for negative values for $Z_i - \tau$ since the logarithm is not defined for negative values. Therefore it follows for the transformed jittered data

$$T^{-1}(Z_i, \tau) \approx \exp(Z_i) + \tau$$

and hence I can state the following corollary.

Corollary 2.10. *The quantile of the transformed jittered data is given as*

$$Q_{T(Z_i,\tau)|x_i}(\tau) = x_i^T\beta_\tau.$$

Proof. The transformation T is almost continuous and thus it holds that

$$Q_{T(Z_i,\tau)|x_i}(\tau) = T\left(Q_{Z_i|x_i}(\tau)\right).$$

In Theorem 2.9 it was shown that

$$Q_{Z_i|x_i}(\tau) = \exp(x_i^T\beta) + \tau,$$

which implies that

$$\begin{aligned} Q_{T(Z_i,\tau)|x_i}(\tau) &= T\left(\exp(x_i^T\beta) + \tau, \tau\right) \\ &= \exp(x_i^T\beta). \end{aligned}$$

□

2.7.2.3 *Applying Quantile Estimation in the Linear Model on the Transformed Jittered Data*

The transformed jittered data

$$Y_i^\star := T(Z_i, \tau)$$

is now continuous and we can apply the quantile estimation in linear models as introduced in Section 2.3. There we estimate β_τ either directly as the minimiser of the loss function (2.7) or as the *maximum likelihood* estimator (2.16). In order to average out the error, which is based on the *jittering*, we apply an *averaged jittering*. That means we jitter our data M times and repeat the estimation of β_τ in each step. In the end we take the averaged estimator

$$\hat{\beta}_\tau = \frac{1}{M} \sum_{m=1}^{M} \hat{\beta}_{\tau,m}.$$

This leads to the quantile estimator of $Y_i^\star$

$$\hat{Q}_{Y_i^\star|x_i}(\tau) = x_i^T \hat{\beta}_\tau, \quad i = 1, 2, \dots, n. \tag{2.26}$$

2.7.2.4 *Back-Transformation and Count Quantile*

From the τ-quantile of $Y_i^\star$ we can calculate the τ-quantile of the observed counts Y_i by the following theorem.

Theorem 2.11. *For a fixed $\tau \in (0, 1)$ the estimator for the τ-quantile of the observed counts Y_{ij} given x_{ij} is given by*

$$\begin{aligned} \hat{Q}_{Y_i|x_i}(\tau) &= \lceil T^{-1}(\hat{Q}_{Z_i|x_i}(\tau)) - 1 \rceil \\ &= \lceil \exp(x_i^T \hat{\beta}_\tau) + \tau - 1 \rceil \end{aligned}$$

for $i = 1, 2, \dots, n$.

Proof. The proof can be found in Machado and Santos Silva [2005] Theorem 2. □

2.7.2.5 *Consistency of the Quantile Estimation of Counts in Linear Mixed Models*

Theorem 2.12. *For a fixed $\tau \in (0, 1)$ the estimator for the τ-quantile of the observed counts Y_i given x_i,*

$$\hat{Q}_{Y_i|x_i}(\tau) \quad i = 1, 2, \dots, n$$

as defined in Theorem 2.11, is consistent.

Proof. The proof can be found in Machado and Santos Silva [2005] Theorem 1 and the Corollary. There the proof of the theorem is based on Pollard [1991] Theorem 1. □

2.7.2.6 *Conclusion of Quantiles of Counts*

In this part I introduced the idea by Machado and Santos Silva [2005] of *jittering* count data in order to apply the linear quantile model. Thus one is able to estimate quantiles of count data by applying the quantile estimation in linear models described in Section 2.3. The quantile estimation works on continuous data, which is why the count data needed to be made continuous by the *jittering* and the transformation. Then we could apply the linear quantile model as in (2.10). After the estimation a back-transformation of the quantile estimators of the transformed jittered data gives the quantiles of the counts.
Furthermore Machado and Santos Silva [2005] showed that the quantile estimation in count data is consistent. This is implied by the consistency of the quantile estimation in linear models, which was stated in Corollary 2.8 of this Chapter.

3 QUANTILE REGRESSION IN LINEAR MIXED MODELS

Similarly as in Chapter 2, starting with a linear model, we are now interested in quantiles in a linear mixed model. For example these models find applications in *Small Area Estimation* (SAE) or on longitudinal data. Therefore I am giving an introduction on linear mixed models in Section 3.1. These models were extended by Geraci and Bottai [2007] for the use in quantile estimation, which is discussed in Sections 3.2 and 3.3. In the following Section 3.4 I show the consistency of the quantile estimator under stated assumptions which is the main contribution of this thesis. Because the main part of this proof is extensive, it has its own Chapter 4. In Section 3.6.1 I discuss the application of quantile estimation in the linear mixed models on count data using a method called *jittering*, which was firstly introduced by Machado and Santos Silva [2005] and already discussed in Section 2.7.2 for linear models. The Chapter ends with Section 3.6.2, where I introduce a method called *Microsimulation via Quantiles* (MvQ), which can be applied to parameter estimation, which goes beyond mean estimation in linear mixed models.

3.1 THE LINEAR MIXED MODEL

Linear mixed models are in common use in statistics. One main application is longitudinal data, where D objects are each observed at different times. Another one is the *Small Area Estimation* (SAE), where D areas each have a within sample size of n_i, $i = 1, 2, \ldots, D$ of individuals or units. Both have in common that dependencies within observations, may they come from the same object or the same area, are caught in a random effect V_i. Applying a linear model without a random effect would violate the independence assumption on the error term. At this point we are interested in the estimation of the mean of the data. This leads to a mean model, the linear mixed model,

$$Y_{ij} = x_{ij}^T \beta + V_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, D; j = 1, 2, \ldots n_i, \tag{3.1}$$

where Y_{ij} is the observation, x_{ij} is a p-dimensional vector of independent variables of the time or individual j in object or area i, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the unknown p-dimensional parameter vector, V_i is the random effect, and ε_{ij} is the individual error. Since I work mainly in the field of SAE, I will name all properties in area and individual terms, keeping in mind that they are exchangeable for other applications of mixed models.

So far there are no distribution assumptions on the error terms V_i and ε_{ij} except that they are centred, thus

$$E[V_i] = 0 \quad \text{and} \quad E[\varepsilon_{ij}] = 0, \quad i = 1, 2, \ldots, D; j = 1, 2, \ldots n_i,$$

and each have a finite variance

$$Var(V_i) = \sigma_V^2 < \infty \quad \text{and} \quad Var(\varepsilon_{ij}) = \sigma_\varepsilon^2 < \infty, \\ i = 1, 2, \ldots, D; j = 1, 2, \ldots n_i.$$

Additionally they are distributed independently of each other. Thus V_{i_1} is distributed independently of $V_{i_2}$for all $i_1 \neq i_2$, $\varepsilon_{i_1 j_1}$ is distributed independently of $\varepsilon_{i_2 j_2}$ for all $(i_1, j_1) \neq (i_2, j_2)$, and V_{i_1} is distributed independently of $\varepsilon_{i_2 j}$ for all $i_1, i_2 = 1, 2, \ldots, D$ and $j = 1, 2, \ldots n_{i_1}$. The sample size in area i is n_i leading to an overall sample size of

$$n = \sum_{i=1}^{D} n_i. \tag{3.2}$$

Of common use is a *normal* assumption on the random effect

$$V_i \overset{iid}{\sim} N(0, \sigma_V^2), \quad i = 1, 2, \ldots, D.$$

Other distributions are possible but have not been used as widely. For example a heavy tailed distribution as e.g. the *Log-normal distribution* or the *t-distribution* should be applied whenever we have extreme value data like income or insurance claims. Also a distribution with a positive support is possible if we are interested in the fastest time a man can run 100 meters for example. A *normal* assumption on the individual error terms is also in common use, especially in SAE approaches:

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2), \quad i = 1, 2, \ldots, D; j = 1, 2, \ldots n_i.$$

We can rewrite model (3.1) in matrix form as follows

$$Y = \mathbf{X}\beta + \mathbf{Z}V + \varepsilon, \tag{3.3}$$

where $Y = (Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1}, Y_{2,1}, \dots, Y_{D,n_D})^T$ is the vector of the observations, the matrix

$$\mathbf{X} := \begin{pmatrix} x_{1,1}^T \\ x_{1,2}^T \\ \vdots \\ x_{1,n_1}^T \\ x_{2,1}^T \\ \vdots \\ x_{D,n_D}^T \end{pmatrix} \tag{3.4}$$

is the design matrix, $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ is the unknown p-dimensional parameter vector, the matrix

$$\mathbf{Z} := \begin{pmatrix} \mathbf{1}_{n_1} & & & \\ & \mathbf{1}_{n_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_D} \end{pmatrix}, \tag{3.5}$$

where $\mathbf{1}_{n_i}$ is an n_i-dimensional vector of ones, is the design matrix for the random vector $V = (V_1, V_2, \dots, V_D)^T$, and $\varepsilon = (\varepsilon_{1,1}, \varepsilon_{1,2}, \dots, \varepsilon_{1,n_1}, \varepsilon_{2,1}, \dots, \varepsilon_{D,n_D})^T$ is the vector of the individual errors. The assumption of *normal distribution*s of the random effect and the individual errors can now be rewritten as

$$V \sim N\left(0_D, \sigma_V^2 I_D\right) \quad \text{and} \quad \varepsilon \sim N\left(0_n, \sigma_\varepsilon^2 I_D\right).$$

This together with the independence of V and ε leads to a *normal distribution* for the observation vector

$$Y \sim N(\mathbf{X}\beta, \Sigma) \tag{3.6}$$

with $\Sigma := \sigma_\varepsilon^2 I_n + \sigma_V^2 \mathbf{Z}\mathbf{Z}^T$. With the assumption of known variances σ_V^2 and σ_ε^2. This leads directly to the best linear unbiased estimator (*BLUE*) by the Gauß-Markov Theorem (cf. Markov [1912], Searle [1971], Henderson [1950])

$$\hat{\beta}(Y) = \left(\mathbf{X}^T \Sigma^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \Sigma^{-1} Y \tag{3.7}$$

and the best linear unbiased predictor (*BLUP*) for the random effect

$$\hat{V}(Y) = \sigma_V^2 \mathbf{Z}^T \Sigma^{-1} \left(Y - \mathbf{X}\hat{\beta}(Y)\right). \tag{3.8}$$

Robinson [1991] gives a good overview on the *BLUP* (3.8) as the predictor for the random effect. It follows the best linear unbiased estimator for Y_{ij} given x_{ij} as

$$\hat{Y}_{ij} = x_{ij}^T \hat{\beta} + \hat{V}_i, \tag{3.9}$$

where $\hat{\beta} = \hat{\beta}(Y)$ from Equation 3.7 and $\hat{V}_i$ is the i^{th} entry of $\hat{V}$ from Equation 3.8.

In general the variance parameters are unknown and need to be estimated first. This leads to the empirical best linear unbiased estimator and predictor (*EBLUE* & *EBLUP*), where the variance parameters in Equations 3.7 and 3.8 are replaced by their estimators $\hat{\sigma}_V^2$ and $\hat{\sigma}_\varepsilon^2$ (Rao [2003], Chapter 6.2.3). This approach is a two-stage method, where the variance parameters are estimated first and then set into the *BLUE* and *BLUP* equations.

Another way of dealing with unknown variance parameters is a *maximum likelihood* approach. From (3.6) follows directly the density and thus the log-likelihood density of the observation Y. The unknown parameters are $\theta = (\sigma_V, \sigma_\varepsilon, \beta^T)^T$. By differentiation of the log-likelihood density and setting this derivative to zero the *maximum likelihood* estimator $\hat{\theta} = (\hat{\sigma}_V, \hat{\sigma}_\varepsilon, \hat{\beta}^T)^T$ can be derived. In a last step the *BLUP* for V is obtained as in Equation 3.8 by replacing the variance components σ_V and σ_ε by their estimates $\hat{\sigma}_V$ and $\hat{\sigma}_\varepsilon$. In the end the best linear unbiased estimator for Y_{ij} given x_{ij} is given as in (3.9). In Searle et al. [1992] can be found a detailed calculation of the variance parameters of this estimator.

This *maximum likelihood* approach is also a two-stage method and is similar to the one we are going to employ for the quantile estimator in mixed models.

3.2 THE LINEAR QUANTILE MIXED MODEL

In Chapter 2 quantile regression in linear models was discussed. However the linear model (2.3) often needs to be improved though. There are scenarios, in which one has to extend this model. Nevertheless linear models serve as a starting point. Koenker [2004] introduced an approach of quantile regression in fixed effects models for the application in longitudinal data. Another example for the necessity of development the linear model (2.3) are dependencies between observations. Linear mixed models with random effects are the right choice in this setting. Furthermore one may even be interested in estimators beyond the mean, which is well developed in mixed models and was introduced in Section 3.1. For quantile estimation in linear mixed models the idea of quantile regression in linear models must be adapted appropriately.

3.2.1 *The Model*

Similarly to the linear quantile model (2.10) without random effect for a fixed $\tau \in (0,1)$, Geraci and Bottai [2007] defined the linear quantile mixed model as follows

$$Q_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T \beta_\tau + V_{\tau,i}, \quad i = 1,2,\dots,D; j = 1,2,\dots n_i, \qquad (3.10)$$

where $Q_{Y_{ij}|x_{ij}}(\tau)$ stands for the conditional τ-quantile of Y_{ij} given x_{ij}. Thus the linear quantile model was extended by adding the random effect $V_{\tau,i}$. The linear quantile mixed model (3.10) only needs to be employed whenever the distribution of the error term in the linear mixed model (3.1) is unknown. For a known error distribution function F_ε, the τ-quantile of Y_{ij} given x_{ij} is then

$$Q_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T \beta + V_i + F_\varepsilon^{-1}(\tau), \quad i = 1,2,\dots,D; j = 1,2,\dots n_i,$$

where β is the same parameter vector as in the linear mixed model (3.1). In practice assuming an unknown distribution of ε leads to more flexibility to the model. Therefore I will proceed under this assumption.

In contrast to the linear mixed model (3.1) the random effect $V_{\tau,i}$ now carries τ in a footnote implying that for different τ the random effect may be different. A further discussion about this approach can be found in Section 3.2.2. In the following I will drop the τ in the subscript for the purpose of simplicity. However the reader should keep in mind the dependency on τ. In matrix form (3.10) can be rewritten as

$$Q_{Y|\mathbf{X}}(\tau) = \mathbf{X}\beta_\tau + \mathbf{Z}V, \qquad (3.11)$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the same matrices as defined in (3.4) and (3.5), respectively. Equivalently to Theorem 2.3 and similar to the approaches by Jung [1996] for median estimation, Koenker and Machado [1999], and Yu and Moyeed [2001], we can state the equivalence of model (3.10) to

$$Y_{ij} = x_{ij}^T \beta_\tau + V_i + \varepsilon_{\tau,ij}, \quad i = 1,2,\dots,D; j = 1,2,\dots n_i \qquad (3.12)$$

with

$$\varepsilon_{\tau,ij} \overset{iid}{\sim} \mathrm{ALD}(0,\sigma,\tau), \quad i = 1,2,\dots,D; j = 1,2,\dots n_i.$$

In matrix form this model can be rewritten as

$$Y = \mathbf{X}\beta_\tau + \mathbf{Z}V + \varepsilon_\tau, \qquad (3.13)$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the same matrices as defined in (3.4) and (3.5), respectively. The error term ε_τ is the vector of the individual error terms $\varepsilon_{\tau,ij}$ in (3.12). Its distribution is an n-dimensional *asymmetric Laplace distribution* as discussed in (2.14)

$$\varepsilon_\tau \overset{iid}{\sim} ALD_n(0_n, \sigma, \tau).$$

Hence the *asymmetric Laplace distribution*, which was discussed in Section 2.4.1, also serves as the distribution of the individual error term $\varepsilon_{\tau,ij}$ here. Again for reasons of simplification I will drop the τ in the subscript of the error term in the following keeping in mind that its distribution is dependent on τ. As in the linear quantile model (2.18) we assume that the scale parameter σ is unknown. Thus it gives a measure of the variance of the individual error term in the linear mixed model (3.1), whose distribution is assumed to be unknown.
Whenever I mention the linear quantile mixed model in the further investigation I refer to the latter model (3.12). Due to the equivalence of the two models, this choice is a matter of taste. I prefer model (3.12) because it has a regular appearance in linear modelling with error terms on the right hand side and the observations on the left hand side. On the other hand model (3.10) carries the error distribution within the quantile expression on the left hand side and there is no direct exposure of the observation Y_{ij} in this model.

3.2.2 *The Dependence of the Random Effect on τ*

At this point I would like to examine whether the random effect V depends on τ. Before I dropped the τ in the subscript for the purpose of simplification in the display. Now let us investigate the dependence here.
In theory the random effect only adds to the linear model. Thus one could interpret the linear quantile mixed model as a combination of the linear quantile model (2.18) and a random effect. Then the quantile may also be rewritten as the quantile of the linear model (cf. (2.10)) added with a random effect

$$Q_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T \beta_\tau + V_i, \quad i = 1, 2, \ldots, D; j = 1, 2, \ldots n_i.$$

Hence all quantiles for an area i have the same distance from the average, V_i, for all τ, and the random effect is independent of τ.

In a simulation study I produced 500 pseudo samples from the linear mixed model with $D = 500$ areas and area sample sizes of $n_i = 10$ individuals in each ($i = 1, 2, \ldots, 500$) with three error scenarios. The model is

$$Y_{ij} = 2 + 0.8x_{ij} + V_i + \varepsilon_{ij},$$

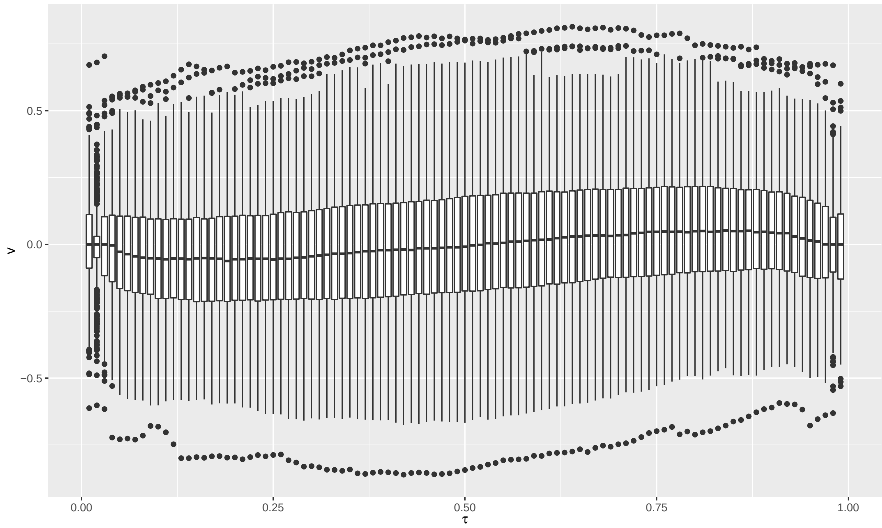

Figure 3.1: Boxplots of predictors of the random effect in the first area of 500 simulation runs for different τ in scenario 1 (symmetric error)

where the independent variables x_{ij} come from a *uniform distribution* on $(0, 1)$

$$x_{ij} \sim U(0,1), \quad i = 1, 2, \dots, 500; j = 1, 2, \dots, 10.$$

The random effect was drawn from a *normal distribution* with zero mean and variance $\sigma_V^2 = 0.3^2$

$$V_i \sim N(0, 0.3^2), \quad i = 1, 2, \dots, 500,$$

and the error term was drawn from a *normal* or a *transformed F distribution* with parameters $d_1 = 20$ and $d_2 = 20$

$$\begin{aligned}
\text{Scenaro 1: } & \varepsilon_{ij} \sim N\left(0, 0.5^2\right) \\
& i = 1, 2, \dots, 500; j = 1, 2, \dots, 10 \\
\text{Scenaro 2: } & \varepsilon_{ij} \sim \sqrt{0.5\frac{20 \cdot 18^2 \cdot 16}{2 \cdot 20^2 \cdot 38}}\left(F(20,20) - \frac{20}{18}\right) \\
& i = 1, 2, \dots, 500; j = 1, 2, \dots, 10 \\
\text{Scenaro 3: } & \varepsilon_{ij} \sim -\sqrt{0.5\frac{20 \cdot 18^2 \cdot 16}{2 \cdot 20^2 \cdot 38}}\left(F(20,20) - \frac{20}{18}\right) \\
& i = 1, 2, \dots, 500; j = 1, 2, \dots, 10
\end{aligned}$$

such that

$$E\left[\varepsilon_{ij}\right] = 0 \quad \text{and} \quad Var(\varepsilon_{ij}) = 0.5^2.$$

I chose these three scenarios such that we have a symmetric and two skewed data from which the estimation of quantiles is more challenging. The first scenario has a symmetric error term. In scenario 2 the error term is positively skewed and in scenario 3 it is negatively

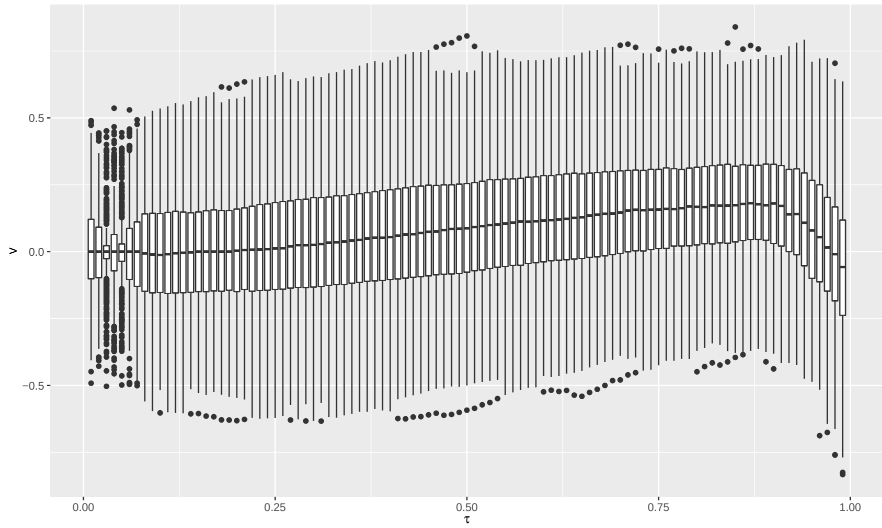

Figure 3.2: Boxplots of predictors of the random effect in the first area of 500 simulation runs for different τ in scenario 2 (positively skewed error)

skewed. I estimated for each sample the parameter $\hat{\beta}_\tau$ and the predictor vector $\hat{V}_\tau$ with the package lqmm by Geraci [2016] in R.
Figures 3.1, 3.2, and 3.3 show boxplots of the predictions of the random effect dependent on τ in the first area of 500 simulation runs for scenario 1, 2, and 3, respectively. In scenario 1, the symmetric setting, the predictors are around zero and have only a small dependence on τ. Around the lower quartile the predictor for the random effect is the lowest while it is the highest for quantiles around the upper quartile. In the skewed scenarios we can see that the prediction for V_1 is further from zero for the quantiles in the tails of the distributions, which in scenario 2 is the upper quantiles. In scenario 3 it is the lower quantiles. This is due to the spread out observations in the tail parts. The quantile estimators become less reliable and this can be seen in the prediction of the random effect.
These were only simulation examples with real quantiles given as

$$Q_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T\beta + V_i + F_\varepsilon^{-1}(\tau), \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots 10$$

with F_ε being the corresponding distribution function in the three scenarios. Thus in theory there should be no dependence of the random effect on τ. However the values for $\hat{V}$ seem to depend on τ because the prediction corrects for less reliability on the data. The estimation of the random effect depends on the distribution of the error term ε which is depending on the fixed τ.
In practise there could be another reason for the dependence of the random effect on τ. In small area estimation it can be interpreted as different random effects of areas for different quantiles. For example the income estimation in the first area may be in the median by $V_{0.5,1}$

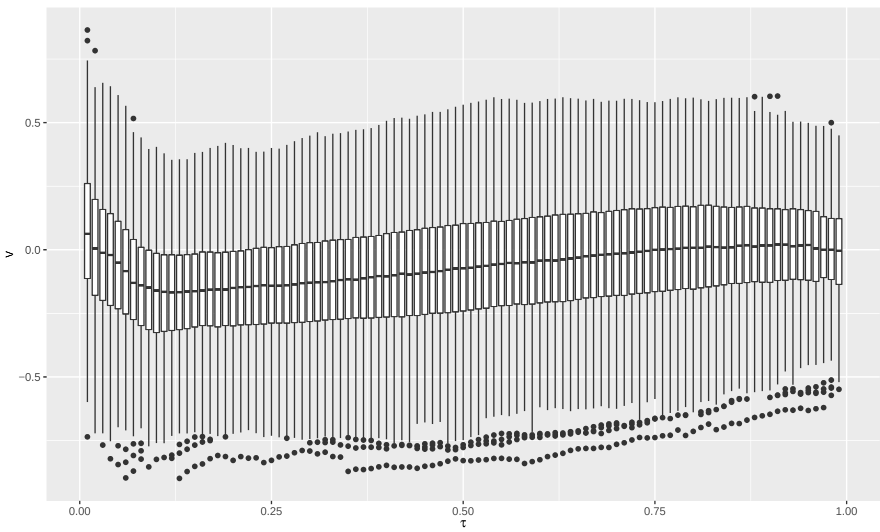

Figure 3.3: Boxplots of predictors of the random effect in the first area of 500 simulation runs for different τ in scenario 3 (negatively skewed error)

higher than the average but in the 95%-quantile is higher by $V_{0.95,1}$ than the average. If $V_{0.5,1} < V_{0.95,1}$ this means that the 95%-quantile is even more distant from the average than the median is. In this area the higher incomes are even larger than the average 95%-quantile by reasons we cannot explain by the independent data $\mathbf{X}_1$. The income is more spread in the upper tail than in other areas. If $V_{0.5,1} > V_{0.95,1}$ this means that the 95%-quantile is less distant from the average than the median is. In this area the higher incomes are lower by reasons we cannot explain by the independent data $\mathbf{X}_1$. The income is less spread in the upper tail than in other areas.
Altogether we can state that the random effect should depend on τ. This makes the model more flexible in terms of the distribution of Y within the areas, which cannot be explained by the independent data $\mathbf{X}$. Nevertheless the footnote on V_τ will be dropped in future appearances for clarity, keeping in mind the dependence on the choice of τ.

3.3 THE QUANTILE ESTIMATION IN LINEAR MIXED MODELS

For the quantile estimator in the linear mixed model we need an estimator for the parameter β_τ and a predictor for the random vector V leading to the conditional quantile estimator for a fixed $\tau \in (0, 1)$

$$\hat{Q}_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T\hat{\beta}_\tau + \hat{V}_i, \quad i = 1, 2, \dots, D; j = 1, 2, \dots n_i. \qquad (3.14)$$

This estimation is fulfilled in two steps, which will be described in the following Sections 3.3.1 and 3.3.2 which is also described in Geraci and Bottai [2007] and Geraci and Bottai [2014] and implemented in the R package lqmm (cf. Geraci [2016]).

3.3.1 *Step 1: Maximum Likelihood Estimation*

From the linear quantile mixed model (3.13) the conditional distribution of Y given V is an *asymmetric Laplace distribution* with location parameter $\mathbf{X}\beta_\tau + \mathbf{Z}V$, scale σ, and skewness τ

$$Y|V \sim \mathrm{ALD}_n\left(\mathbf{X}\beta_\tau + \mathbf{Z}V, \sigma, \tau\right).$$

Thus the joint distribution of the observation vector Y and the random effect vector V is given as the convolution

$$(Y, V) \sim \mathrm{ALD}_n\left(\mathbf{X}\beta_\tau + \mathbf{Z}V, \sigma, \tau\right) \times N_D\left(0_D, \sigma_V^2 I_D\right).$$

It follows that the density of the joint distribution is given as

$$f_{(Y,V)}(y, v) = f_{\mathrm{ALD}_n(\mathbf{X}\beta_\tau + \mathbf{Z}V, \sigma, \tau)}(y|v) \cdot f_{N_D\left(0_D, \sigma_V^2 I_D\right)}(v).$$

This can be simplified as in (2.14) to the joint distribution density

$$f_{(Y,V)}(y, v) = \prod_{i=1}^{D}\left(\prod_{j=1}^{n_i} f_{\mathrm{ALD}\left(x_{ij}^T\beta_\tau + v_i, \sigma, \tau\right)}(y_{ij}|v_i)\right) f_{N\left(0, \sigma_V^2\right)}(v_i). \tag{3.15}$$

The density and thus the distribution of the observation vector Y is then given as the marginal density of the joint density in (3.15)

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}^D} \prod_{i=1}^{D}\left(\prod_{j=1}^{n_i} f_{\mathrm{ALD}\left(x_{ij}^T\beta_\tau + v_i, \sigma, \tau\right)}(y_{ij}|v_i)\right) f_{N\left(0, \sigma_V^2\right)}(v_i) dv \\ &\overset{(\star)}{=} \prod_{i=1}^{D} \int_{\mathbb{R}} \left(\prod_{j=1}^{n_i} f_{\mathrm{ALD}\left(x_{ij}^T\beta_\tau + v_i, \sigma, \tau\right)}(y_{ij}|v_i)\right) f_{N\left(0, \sigma_V^2\right)}(v_i) dv_i, \end{aligned} \tag{3.16}$$

where $(\star)$ follows by application of the *Theorem of Fubini* (cf. Klenke [2013], Satz 14.16). A closed form solution of this integral is not calculable. Thus (3.16) is the simplified expression of the density of the observation Y. The unknown parameters are σ_V, σ, and β_τ. From 3.16 we can derive the log-likelihood density $\ell(\theta|y) = \log f_Y(y)$ and find the *maximum likelihood* estimator

$$\hat{\theta} := \arg\max_{\theta \in \Theta} \ell(\theta|y).$$

for $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T \in \Theta := \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^p$ as the roots of the equations

$$\frac{\partial}{\partial\theta}\ell(\theta|y) \overset{!}{=} 0_{p+2}. \tag{3.17}$$

Since there is no analytical solution to (3.17), numerical approaches are needed. Geraci and Bottai [2007] first introduced an *EM algorithm* (cf. McLachlan and Krishnan [2008]). Later Geraci and Bottai [2014] made use of a *Gaussian quadrature* (cf. Pinheiro and Chao [2006]). The latter procedure, which is also implemented in their R package lqmm (cf. Geraci [2016]), is faster and more stable than the *EM algorithm*. The user is able to assume different distributions on the random effect and the number of knots in the *Gaussian quadrature*. As a result the *maximum likelihood* estimator $\hat{\theta} = (\hat{\sigma}_V, \hat{\sigma}, \hat{\beta}_\tau^T)^T$ can be calculated.
The existence and consistency of this *maximum likelihood* estimation is proven in Chapter 4.

3.3.2 *Step 2: Prediction of Random Effect*

In a second step a prediction for the random effect is calculated using the *maximum likelihood* estimator $\hat{\theta} = (\hat{\sigma}_V, \hat{\sigma}, \hat{\beta}_\tau^T)^T$ from Step 1 introduced in Section 3.3.1. As in linear mixed models (3.1) Geraci and Bottai [2014] stated that the predictor for the random effect can be written in the linear quantile mixed model (3.12) as

$$\hat{V}(Y) = \hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(Y - \mathbf{X}\hat{\beta}_\tau - \hat{E}\left[\varepsilon\right]\right) \tag{3.18}$$

with the estimated covariance matrix of Y

$$\hat{\Sigma} = \hat{\sigma}_V^2 \mathbf{Z}\mathbf{Z}^T + \widehat{Var}(\varepsilon)$$

and the estimated expected value and variance of ε are

$$\hat{E}\left[\varepsilon\right] = \frac{\hat{\sigma}(1-2\tau)}{\tau(1-\tau)}\mathbf{1}_n \quad \text{and}$$
$$\widehat{Var}(\varepsilon) = \frac{\hat{\sigma}^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2} I_n.$$

These are the expected value and the variance of an n-dimensional *asymmetric Laplace distribution* with parameters $\mu = 0$, $\hat{\sigma}$, and τ (cf. Corollary 2.2). Note that the estimated covariance matrix can also be rewritten as $\hat{\Sigma} = \hat{\sigma}_V^2 \mathbf{Z}\mathbf{Z}^T + \widehat{Var}(\varepsilon_{1,1}) I_n$ with $\widehat{Var}(\varepsilon_{1,1}) = \frac{\hat{\sigma}^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2}$.
The approach is based on the best linear prediction (cf. Ruppert et al. [2003], Chapter 4.3.1).
As a result the quantile estimator given in (3.14) can be calculated by inserting $\hat{\beta}_\tau$ from the *maximum likelihood* estimation in Step 1 and $\hat{V}_i$ given in Equation 3.18.

3.4 THE CONSISTENCY OF QUANTILE ESTIMATION IN THE LINEAR MIXED MODEL

In this Section I am going to show the consistency of the quantile estimator in linear mixed models as introduced in (3.14).

3.4.1 *The Consistency of the Quantile Estimator*

For the term consistency I am going to use the Definition 2.4 which was stated in Chapter 2, when I discussed the consistency of the quantile estimation in linear models. Similar to the two-step approach in the estimation, the consistency is separately shown for both steps. Since the *maximum likelihood* estimation in the first step is quite complicated and the proof is going to be of analytical kind, this part is sourced out in a whole Chapter 4. Nevertheless I am giving a short version of this part in the proof of the following theorem. Before I state the consistency let me assume the following properties:

(B1) Let the sample size $n \to \infty$ with $D \to \infty$ and $n_i \to \infty$ for all $i = 1, 2, \ldots, D$.

(B2) The true value θ^0 is in the interior of $\Theta = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^p$.

(B3) $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} \to c_0$ for $n \to \infty$, where c_0 is a p-dimensional vector.

(B4) $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}x_{ij}^T \to C_1$ for $n \to \infty$, where C_1 is a positive definite $p \times p$-matrix.

(B5) $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} f_{Y_{ij}}(x_{ij}\beta_\tau + v_i)x_{ij}x_{ij}^T \to C_2$, where C_2 is a positive definite $p \times p$-matrix.

(B6) The conditional expectations $E[\varepsilon_{i,j}|Y]$ and $E[V_i|Y]$ are independent for all $i = 1, \ldots, D; j = 1, \ldots, n_i$.

3.4.1.1 *The Meaning of the Assumptions*

Before I state the main theorem of this thesis let me discuss the assumptions needed there. The first two assumptions (B1) and (B2) can also be found in the consistency proof in the mean estimation in linear mixed models in Pinheiro [1994]. Especially (B1) is necessary in mixed models in order to obtain the consistency of the variance estimator of the random effect $\hat{\sigma}_V^2$. So the overall sample size n must increase while, in small area terms, the number of areas D and the number of observations in all areas n_i grow.
On the other hand (B3) and (B4) are of technical character and comparable to Assumption A3.1.7 in Pinheiro [1994]. They are later needed to show the positive definiteness of the inverse covariance matrix B (see Lemma 4.15) and within the proof of the Weiss' Assumptions

(A1) and (A2) stated before Theorem 4.1. As we assume a random design on the covariates x_{ij} the vector c_0 in (B3) has the character of a mean $E[X_{1,1}]$. Here $X_{1,1}$ must be understood as the represent of a covariate vector.

The assumption (B5) on the design is needed in order to apply the *Law of Large Numbers* at various points within the consistency proof. Pinheiro [1994] does not apply this law and therefore he has no comparable assumption. However a similar assumption can be found in the consistency proof of quantiles in linear models (e.g. as assumption A2(ii) in Koenker [2005]). Therefore it is a standard assumption in quantile regression. This assumption is comparable to the regularity condition

$$\sum_{i,j} x_{ij} x_{ij}^T \to D \quad \text{where D is a positive definite matrix}$$

in mean regression.

The application of the *Law of Large Numbers* $\frac{1}{n}\sum_{i,j}(Y_{ij} - E\left[Y_{ij}\right]) \to 0$ needs the expectation $E\left[Y_{ij}\right]$ to be independent from i and j which is in general not the case in linear regression. However by assuming (B5) on the design matrix X the *Law of Large Numbers* is also applicable here.

The last assumption (B6) is also of technical character. As assumed in the linear quantile mixed model (3.12) the random effect V_i and the error term ε_{ij} are already independent. Within the consistency proof there will be the conditional expectations with respect to the observation Y, $E[\varepsilon_{1,1}|Y]$ and $E[V_1|Y]$. As in the discussion of (B3) $\varepsilon_{1,1}$ and V_1 must be understood as the represents of the error terms and the random effect, respectively. The conditional expectation $E[\varepsilon_{1,1}|Y]$ and $E[V_1|Y]$ are again random variables and their independence does not naturally follow from the independence of $\varepsilon_{1,1}$ and V_1. Therefore I must state this assumption at this point.

Pinheiro [1994] stated further assumptions A3.1.1 to A3.1.4. in his consistency proof. These are assumptions Miller [1977] stated in his application of the Weiss' Theorem (see Theorem 4.1). In my case there is no need for these assumptions since I only assume one level of random effect and the number of observations n large enough while the length of the covariates x_{ij}, p, is fixed.

In comparison to the quantile estimation in linear models as discussed in Chapter 2 there are more conditions in the mixed model. Theorem 2.7 states the asymptotic normality of the estimator $\hat{\beta}_\tau$. The assumption (a) there is formulated within Theorem 3.1 and not as a condition here. (B4) and (B5) are similar to (b) and (c) in the theorem. Condition (d) from Theorem 2.7 is comparable to (B3). (B1) is only needed in the mixed model because it addresses the sample size D while (B2) is hidden within Theorem 2.7 but can be found in the more general Theorem 2.6 (cf. condition (a) there). (B6) is, as mentioned, of technical character and can not be found in the linear quantile model case.

Theorem 3.1. *Let (B1) to (B6) hold. For a fixed* $\tau \in (0,1)$ *the quantile estimator*

$$\hat{Q}_{Y_{ij}|x_{ij}}(\tau) = x_{ij}^T \hat{\beta}_\tau + \hat{V}_i, \quad i = 1,2,\dots,D; j = 1,2,\dots n_i.$$

is consistent.

Proof. As mentioned before this proof is divided into two parts. First I am proving the consistency of the *maximum likelihood* estimator from Step 1 as described in Section 3.3.1. In a second step I am showing the consistency of the prediction of the random effect as described in Section 3.3.2.
Step 1: The consistency of the parameter estimation of $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T$ is shown in detail in Chapter 4. There the log-likelihood density is calculated and twice differentiated in order to apply a theorem for this non-standard *maximum likelihood* estimation. It is not standard, because all observations from the same area i Y_{i,j_1} and Y_{i,j_2}, $j_1 \neq j_2$, are dependent due to the joint random effect in the linear quantile mixed model (3.12). Thus the standard *maximum likelihood* approach, as stated in Theorem 2.6, is not applicable. The theorem applied in this case is the *Weiss' Theorem* –Theorem 4.1 (cf. Weiss [1971] and Weiss [1973]). There it is stated that under two assumptions the parameter estimators exist and are *asymptotically normal* with a rate $K(n)$ and a covariance matrix $B^{-1}(\theta^0)$.
As a preparation for the proof of the two assumptions, I have to calculate the second derivatives of the log-likelihood density

$$\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|y), \quad \iota_1, \iota_2 = 1,2,\dots,p+2.$$

This has to be executed for all $p+2$ unknown parameters. Remember β_τ is a p-dimensional vector. By treating β_τ as one parameter and the ability to interchange directions of differentiation – see Section 4.1.3.4.1 –, I am able to decrease the number of second derivatives from $(p+2)^2$ to six. Thus the second derivatives in these six cases are given in Lemmata 4.9, 4.10, 4.11, 4.12, 4.13, and 4.14.
The first assumption of the *Weiss' Theorem* is the convergence in probability of the second derivatives divided by the convergence rate $K_{\iota_1}(n)K_{\iota_2}(n)$ at the true value θ^0 to a continuous function $B_{\iota_1,\iota_2}(\theta^0)$, which builds a positive definite matrix

$$-\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y)\Big|_{\theta^0} \xrightarrow{P} B_{\iota_1,\iota_2}(\theta^0).$$

I show this convergence in the proof of Lemma 4.15. In this Lemma the matrix $B(\theta^0)$ is given and it turns out to be the inverse of the

asymptotic covariance matrix of the estimator $\hat{\theta}$. In Section 4.2.1.3 I show the continuity of $B_{\iota_1,\iota_2}(\theta)$ in θ^0 and in Section 4.2.1.4 I prove the positive definiteness of the matrix $B(\theta^0)$.
The second assumption of the *Weiss' Theorem* states the speed of the convergence in the first assumption. The proof of it is a matter of construction. It is shown in detail for all six cases in Section 4.2.2.
After showing both assumptions the application of the *Weiss' Theorem* leads to the *asymptotic normality* (cf. Chapter 9.3 of van der Vaart [2007]) of the parameter estimator

$$\begin{pmatrix} \sqrt{D}(\hat{\sigma}_V(n) - \sigma_V^0) \\ \sqrt{n}(\hat{\sigma}(n) - \sigma^0) \\ \sqrt{n}(\hat{\beta}_\tau(n) - \beta_\tau^0) \end{pmatrix} \overset{\mathcal{D}}{\to} N(0, B^{-1}(\theta^0))$$

with the rates $\sqrt{D}$ for $\hat{\sigma}_V(n)$ and $\sqrt{n}$ for the other parameters.
In Section 4.3.1 I show that this *asymptotic normality* implies the consistency of the parameter estimator $\hat{\theta} = (\hat{\sigma}_V, \hat{\sigma}, \hat{\beta}_\tau^T)^T$.
Step 2: In order to show the consistency of $\hat{Q}_{Y_{ij}|x_{ij}}(\tau)$ I need to show the consistency of $\hat{\beta}_\tau$ and $\hat{V}_i$. The former was shown in the first step of this proof or in detail in Chapter 4. Hence the proof of consistency of the predictor $\hat{V}$ is remaining.
Let me observe the distance from $\hat{V}$ as stated in (3.18) to the true random vector V.

$$\left|\hat{V} - V\right| = \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(Y - \mathbf{X}\hat{\beta}_\tau - \hat{E}\left[\varepsilon\right]\right) - V\right|.$$

With the linear quantile mixed model (3.13) it follows

$$\begin{aligned} \left|\hat{V} - V\right| &= \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\mathbf{X}\beta_\tau + \mathbf{Z}V + \varepsilon - \mathbf{X}\hat{\beta}_\tau - \hat{E}\left[\varepsilon\right]\right) - V\right| \\ &= \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\mathbf{X}(\beta_\tau - \hat{\beta}_\tau) + \varepsilon - \hat{E}\left[\varepsilon\right]\right)\right. \\ &\qquad\qquad \left. + \left(\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Z} - I_D\right) V\right| \\ &\overset{(\star)}{\leqslant} \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{X}(\beta_\tau - \hat{\beta}_\tau)\right| + \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\varepsilon - \hat{E}\left[\varepsilon\right]\right)\right| \\ &\quad + \left|\left(\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Z} - I_D\right) V\right|, \end{aligned} \tag{3.19}$$

where $(\star)$ follows from the *triangle inequality*. The convergences $\hat{\sigma}_V \overset{n\to\infty}{\to} \sigma_V$ and $\hat{\sigma} \overset{n\to\infty}{\to} \sigma$ (as shown in Step 1) imply for the covariance matrix estimator

$$\hat{\Sigma} \overset{n\to\infty}{\to} \Sigma. \tag{3.20}$$

Let $\epsilon > 0$ be arbitrary and fixed. From Step 1 I know that $\hat{\beta}_\tau$ is consistent, $\hat{\sigma}_V \overset{n\to\infty}{\to} \sigma_V$, and (3.20) holds. Thus for the first summand of (3.19) it holds

$$P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{X}(\beta_\tau - \hat{\beta}_\tau)\right| \geqslant \frac{\epsilon}{3}\right) \overset{n\to\infty}{\to} 0. \tag{3.21}$$

By the *triangle inequality* the second summand of (3.19) can be further calculated to

$$\begin{aligned}
&\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\varepsilon - \hat{E}[\varepsilon]\right)\right| \\
&\quad \leqslant \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\varepsilon - \frac{\sigma(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n\right)\right| \\
&\qquad + \left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\frac{\sigma(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n - \hat{E}[\varepsilon]\right)\right|.
\end{aligned}$$

The convergences $\hat{\sigma}_V \overset{n\to\infty}{\to} \sigma_V$, (3.20), and the *Law of Large Numbers* imply for the first summand

$$P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\varepsilon - \frac{\sigma(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n\right)\right| \geqslant \frac{\epsilon}{6}\right) \overset{n\to\infty}{\to} 0.$$

The estimated expectation of ε is given as

$$\hat{E}[\varepsilon] = \frac{\hat{\sigma}(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n$$

and thus for the second summand it holds by $\hat{\sigma}_V \overset{n\to\infty}{\to} \sigma_V$, (3.20), and the consistency of $\hat{\sigma}$ shown in Step 1

$$\begin{aligned}
&P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\frac{\sigma(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n - \frac{\hat{\sigma}(1-2\tau)}{\tau(1-\tau)} \mathbf{1}_n\right)\right| \geqslant \frac{\epsilon}{6}\right) \\
&\quad = P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \frac{1-2\tau}{\tau(1-\tau)} (\sigma - \hat{\sigma}) \mathbf{1}_n\right| \geqslant \frac{\epsilon}{6}\right) \overset{n\to\infty}{\to} 0.
\end{aligned}$$

Altogether this leads to

$$P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \left(\varepsilon - \hat{E}[\varepsilon]\right)\right| \geqslant \frac{\epsilon}{3}\right) \overset{n\to\infty}{\to} 0. \tag{3.22}$$

The matrix $\hat{\Sigma} = \hat{\sigma}_V^2 \mathbf{Z}\mathbf{Z}^T + \widehat{Var}(\varepsilon_{1,1}) I_n$ is a block diagonal matrix with D blocks $\hat{\Sigma}_i$ of size $n_i \times n_i$, $i = 1, 2, \ldots, D$. The blocks can be written as

$$\hat{\Sigma}_i = \widehat{Var}(\varepsilon_{1,1}) I_{n_i} + \hat{\sigma}_V^2 \mathbf{1}_{n_i \times n_i}, \quad i = 1, 2, \ldots, D.$$

The inverse matrix of a block diagonal matrix is given as the block diagonal matrix with the inverse block matrices:

$$\hat{\Sigma}^{-1} = \begin{pmatrix} \hat{\Sigma}_1 & & & \\ & \hat{\Sigma}_2 & & \\ & & \ddots & \\ & & & \hat{\Sigma}_D \end{pmatrix}^{-1} = \begin{pmatrix} \hat{\Sigma}_1^{-1} & & & \\ & \hat{\Sigma}_2^{-1} & & \\ & & \ddots & \\ & & & \hat{\Sigma}_D^{-1} \end{pmatrix}.$$

By multiplying $\hat{\Sigma}^{-1}$ with $\mathbf{Z}^T$ from the left hand side and with $\mathbf{Z}$ from the right hand side it leads to a $D \times D$-dimensional matrix with the sums of the entries of $\hat{\Sigma}_i^{-1}$ on the diagonal

$$\mathbf{Z}^T\hat{\Sigma}^{-1}\mathbf{Z} = \begin{pmatrix} \sum_{k,l=1}^{n_1} \left(\hat{\Sigma}_1^{-1}\right)_{kl} & & & \\ & \sum_{k,l=1}^{n_2} \left(\hat{\Sigma}_2^{-1}\right)_{kl} & & \\ & & \ddots & \\ & & & \sum_{k,l=1}^{n_D} \left(\hat{\Sigma}_D^{-1}\right)_{kl} \end{pmatrix}.$$

By applying Lemmata A.1 and A.2 on each $\hat{\Sigma}_i$, $i = 1, 2, \ldots, D$, with $a = \widehat{Var}(\varepsilon_{1,1})$, $b = \hat{\sigma}_V^2$, and $n = n_i$ I get

$$\sum_{k,l=1}^{n_i} \left(\hat{\Sigma}_i^{-1}\right)_{kl} = \frac{n_i}{\widehat{Var}(\varepsilon_{1,1}) + n_i\hat{\sigma}_V^2} \overset{n_i \to \infty}{\to} \frac{1}{\hat{\sigma}_V^2}, \quad i = 1, 2, \ldots, D.$$

This implies for all $n_i \to \infty$, $i = 1, 2, \ldots, D$, which holds under (B1), that

$$\mathbf{Z}^T\hat{\Sigma}^{-1}\mathbf{Z} \to \frac{1}{\hat{\sigma}_V^2} I_D.$$

Thus the third summand in (3.19) converges to zero as all $n_i \to \infty$, $i = 1, 2, \ldots, D$,

$$\left|\left(\hat{\sigma}_V^2\mathbf{Z}^T\hat{\Sigma}^{-1}\mathbf{Z} - I_D\right) V\right| \to \left|\left(\hat{\sigma}_V^2 \frac{1}{\hat{\sigma}_V^2} I_D - I_D\right) V\right| = 0,$$

which implies

$$P\left(\left|\left(\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Z} - I_D\right) V\right| < \frac{\epsilon}{3}\right) \overset{n\to\infty}{\to} 0. \tag{3.23}$$

Combining (3.19), (3.21), (3.22), and (3.23) leads to

$$\begin{aligned}
&P\left(\left|\hat{V}(Y) - V\right| \geqslant \epsilon\right) \\
&\quad \leqslant P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{X}(\beta_\tau - \hat{\beta}_\tau)\right| \geqslant \frac{\epsilon}{3}\right) + P\left(\left|\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1}\left(\varepsilon - \hat{E}[\varepsilon]\right)\right| \geqslant \frac{\epsilon}{3}\right) \\
&\quad + P\left(\left|\left(\hat{\sigma}_V^2 \mathbf{Z}^T \hat{\Sigma}^{-1} \mathbf{Z} - I_D\right) V\right| \geqslant \frac{\epsilon}{3}\right) \overset{n\to\infty}{\to} 0
\end{aligned}$$

and thus the consistency of the random effect predictor $\hat{V}(Y)$, as defined in (3.18), is shown.
Conclusion: In Step 1 and in Step 2 of this proof I have shown that the *maximum likelihood* estimator $\hat{\beta}_\tau$ and the predictor $\hat{V}$ are both consistent. For the quantile estimator in the linear mixed model it follows for an arbitrary but fixed $\epsilon > 0$

$$\begin{aligned}
&P\left(\left|\hat{Q}_{Y_{ij}|x_{ij}}(\tau) - Q_{Y_{ij}|x_{ij}}(\tau)\right| \geqslant \epsilon\right) \\
&= P\left(\left|x_{ij}^T \hat{\beta}_\tau + \hat{V}_i - x_{ij}^T \beta_\tau - V_i\right| \geqslant \epsilon\right) \\
&= P\left(\left|x_{ij}^T\left(\hat{\beta}_\tau - \beta_\tau\right) + \hat{V}_i - V_i\right| \geqslant \epsilon\right) \\
&\overset{(\star)}{\leqslant} P\left(\left|x_{ij}^T\left(\hat{\beta}_\tau - \beta_\tau\right)\right| \geqslant \frac{\epsilon}{2}\right) + P\left(\left|\hat{V}_i - V_i\right| \geqslant \frac{\epsilon}{2}\right),
\end{aligned} \tag{3.24}$$

where $(\star)$ holds by the *triangle inequality* and both probabilities converge to zero as $n \to \infty$, because of the before shown consistency of $\hat{\beta}_\tau$ and $\hat{V}$. □

As seen in the second step of the previous proof the consistency of the parameter estimation in the *maximum likelihood* step implies the consistency of the whole quantile estimation in the linear mixed model. The prediction of the random effect is of same form as in Equation 3.18 in the linear mixed model (3.1) with *normal* error terms. There it holds that $\hat{E}[\varepsilon] = 0$ – see (3.8). The consistency of $\hat{V}$ can be shown in this model with the same argumentation as in the previous proof in Step 2 under the assumption of consistent estimators for σ_V, σ_ε, and β. The consistency in the *maximum likelihood* step was shown by Pinheiro [1994] leading to a consistent mean estimator (3.9) in the linear mixed model.

3.4.2 *Simulation Study of the Quantile Estimation in Linear Mixed Models*

In order to display the stated consistency in Theorem 3.1 let us simulate data from the linear quantile mixed model (3.12) and observe the quantile estimation with increasing sample sizes.

In a simulation study I produced each 500 pseudo populations with 500 areas with $N_i = 200$ individuals each ($i = 1, 2, \ldots, 500$) and thus $N = \sum_{i=1}^{D} N_i = 100000$. The model used for $\tau = 0.6$ and $\tau = 0.8$ is

$$Y_{ij} = 2 + 0.8x_{ij} + V_i + \varepsilon_{ij},$$

where the independent variables x_{ij} come from a *uniform distribution* on $(0, 1)$

$$x_{ij} \sim U(0, 1), \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots, 200.$$

The random effect was drawn from a *normal distribution* with zero mean and variance $\sigma_V^2 = 0.3^2$

$$V_i \sim N(0, 0.3^2), \quad i = 1, 2, \ldots, 500,$$

and the error term was drawn from an *asymmetric Laplace distribution* with scale parameter $\sigma = 0.5$ and $\tau = 0.6$ or $\tau = 0.8$

$$\varepsilon_{ij} \sim ALD(0, 0.5, \tau), \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots, 200.$$

In different scenarios I sampled the d first areas from the $D = 500$ with $d = 10, 50, 100, 200, 350, 500$ and sampled from each sampled area n observations with $n_i = 10, 20, 30, 50, 100, 200$. In each population Pop ($\text{Pop} = 1, 2, \ldots, 500$) I estimated for each sample the parameter vector $\hat{\beta}_\tau^{\text{Pop},d,n_i}$ and the predictor vector $\hat{V}_\tau^{\text{Pop},d,n_i}$ with the package lqmm by Geraci [2016] in R leading to quantile estimators

$$\begin{aligned}\hat{Q}^{\text{Pop},d,n_i}_{Y_{ij}|x_{ij}}(\tau) = \hat{\beta}^{\text{Pop},d,n_i}_{\tau,1} + \hat{\beta}^{\text{Pop},d,n_i}_{\tau,2} x_{ij} + \hat{V}^{\text{Pop},d,n}_{i}&,\\ i = 1, 2, \ldots, 500; j = 1, 2, \ldots 200&;\\ d = 10, 50, 100, 200, 350, 500&;\\ n_i = 10, 20, 30, 50, 100, 200&.\end{aligned}$$

Note that $\hat{V}_i^{\text{Pop},d,n}$ is set to zero for non-sampled areas. Since I used the same β_τ for the two quantiles it holds $\beta_{0.6} = \beta_{0.9} = (2, 0.8)^T$. Thus the real quantiles are given as

$$Q^{\text{Pop}}_{Y_{ij}|x_{ij}}(\tau) = 2 + 0.8x_{ij} + \nu_i^{\text{Pop}}, \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots 200,$$

where ν_i^{Pop} was simulated and fixed. In each scenario and population I calculated the *mean absolute deviation* (MAD) for every population and the different sample sizes

$$\text{MAD}^{\text{Pop},d,n_i} = \frac{1}{N} \sum_{i=1}^{500} \sum_{j=1}^{200} \left| \hat{Q}^{\text{Pop},d,n_i}_{Y_{ij}|x_{ij}}(\tau) - Q^{\text{Pop}}_{Y_{ij}|x_{ij}}(\tau) \right|. \tag{3.25}$$

This measure shall represent the absolute distance of $\hat{Q}_{Y_{ij}|x_{ij}}(\tau)$ and

$\tau = 0.6$	$d = 10$	$d = 50$	$d = 100$	$d = 200$	$d = 350$	$d = 500$
$n = 10$	0.2888 0.0496	0.2449 0.01180	0.2366 0.0092	0.2271 0.0079	0.2148 0.0075	0.2035 0.0072
$n = 20$	0.2743 0.0426	0.2393 0.0100	0.2301 0.0083	0.2164 0.0076	0.1976 0.0068	0.1790 0.0060
$n = 30$	0.2647 0.0342	0.2367 0.0096	0.2263 0.0084	0.2093 0.0074	0.1855 0.0064	0.1619 0.0055
$n = 50$	0.2629 0.0374	0.2337 0.0095	0.2212 0.0082	0.1998 0.0072	0.1688 0.0057	0.1386 0.0045
$n = 100$	0.2631 0.0438	0.2317 0.0113	0.2154 0.0086	0.1877 0.0069	0.1472 0.0052	0.1072 0.0036
$n = 200$	0.2636 0.0423	0.2320 0.0159	0.2110 0.0092	0.1772 0.0068	0.1284 0.0049	0.0803 0.0025
$\tau = 0.8$	$d = 10$	$d = 50$	$d = 100$	$d = 200$	$d = 350$	$d = 500$
$n = 10$	0.3051 0.0610	0.2508 0.0154	0.2433 0.0095	0.2369 0.0087	0.2310 0.0082	0.2259 0.0079
$n = 20$	0.2755 0.0369	0.2449 0.0119	0.2386 0.0093	0.2328 0.0093	0.2270 0.0118	0.2221 0.0156
$n = 30$	0.2693 0.0312	0.2420 0.0106	0.2355 0.0087	0.2280 0.0093	0.2177 0.0128	0.2083 0.0161
$n = 50$	0.2709 0.1294	0.2385 0.0098	0.2313 0.0085	0.2187 0.0083	0.2018 0.0084	0.1850 0.0074
$n = 100$	0.2942 0.2321	0.2350 0.0091	0.2248 0.0081	0.2064 0.0071	0.1809 0.0061	0.1560 0.0051
$n = 200$	0.4830 0.5058	0.2326 0.0102	0.2190 0.0083	0.1560 0.0079	0.1592 0.0054	0.1247 0.0063

Table 3.1: Empirical means and standard deviations of MAD^{Pop,d,n_i} in 500 estimations dependent on $d = 10, 50, 100, 200, 350, 500$ and $n = 10, 20, 30, 50, 100, 200$ for $\tau = 0.6$ and $\tau = 0.8$

$Q_{Y_{ij}|x_{ij}}(\tau)$ in (3.24), which is shown to converge in probability to zero. Table 3.1 shows the empirical means and standard deviations of the 500 values of MAD^{Pop,d,n_i} for each population dependent on d and n_i. There the convergence of the mean and the standard deviation to zero for increasing d and n_i is visible for both quantile estimators. The estimation for $\tau = 0.6$ outperforms the estimation for $\tau = 0.8$, due to the better location of the 0.6-quantile in the middle of the data. An interesting observation is the column of $d = 10$ for $\tau = 0.8$. There the estimation seems to become worse with increasing n. In the same column for $\tau = 0.6$ the estimations also seems to stagnate. This can be explained by the higher weight of the observed areas with increasing n_i. This may lead to a worse estimation of the intercept $\beta_{\tau,1}$ if the sampled areas have similar random effects (mainly negative or positive). Thus the whole quantile estimator changes for the worse in these populations. This is also explained by the increasing standard deviation, especially for $\tau = 0.8$. However the consistency was shown for increasing d and n_i. It can be seen on the diagonals of the Tabular 3.1. Hence this worsen in estimation for a small d is no contradiction to the theory.

3.5 THE MEAN SQUARED ERROR OF QUANTILE ESTIMATION IN LINEAR MIXED MODELS

In this Section I am going to examine the *mean squared error* of the quantile estimator in linear mixed models. First let me define the *mean squared error* or *MSE* of an estimator.

Definition 3.2. *For an observation vector* Y *the mean squared error of an estimator* $\hat{S}(Y)$ *is defined as*

$$\mathrm{MSE}(\hat{S}(Y)) := E\left[\left(\hat{S}(Y) - S^0\right)^2\right],$$

where S^0 *is the true but unknown value of interest.*

The *mean squared error* may be decomposed into the *bias* and the variance of the estimator as stated and proven in the following theorem.

Theorem 3.3. *The mean squared error of an estimator* $\hat{S}(Y)$ *can be decomposed as follows*

$$\mathrm{MSE}(\hat{S}(Y)) = \mathrm{Bias}^2(\hat{S}(Y)) + \mathrm{Var}(\hat{S}(Y)),$$

where the bias is defined as

$$\mathrm{Bias}(S(Y)) := E\left[\hat{S}(Y) - S^0\right] = E\left[\hat{S}(Y)\right] - S^0$$

and the variance is given as

$$\mathrm{Var}(\hat{S}(Y)) = E\left[\left(\hat{S}(Y) - E\left[\hat{S}(Y)\right]\right)^2\right].$$

Proof.

$$\begin{aligned}
\mathrm{MSE}(\hat{S}(Y)) &:= E\left[\left(\hat{S}(Y) - S^0\right)^2\right] \\
&= E\left[\hat{S}(Y)^2\right] - 2E\left[\hat{S}(Y)\right]S^0 + {S^0}^2 \\
&= E\left[\hat{S}(Y)^2\right] - E^2\left[\hat{S}(Y)\right] + E^2\left[\hat{S}(Y)\right] - 2E\left[\hat{S}(Y)\right]S^0 + {S^0}^2 \\
&= \mathrm{Var}\left(\hat{S}(Y)\right) + \left(E\left[\hat{S}(Y)\right] - S^0\right)^2 \\
&= \mathrm{Var}\left(\hat{S}(Y)\right) + \mathrm{Bias}^2(\hat{S}(Y))
\end{aligned}$$

□

The expression in Theorem 3.3 works only for one-dimensional estimators $\hat{S}(Y)$. Sometimes estimators are of greater dimension and then the following generalisation holds.

Remark 3.4. *If $\hat{S}(Y)$ is a vector estimator the decomposition of its mean squared error is given as*

$$MSE(\hat{S}(Y)) = \left\|\mathrm{Bias}(\hat{S}(Y))\right\|^2 + \mathrm{trace}(\mathrm{Var}(\hat{S}(Y))),$$

where the bias is defined as

$$\mathrm{Bias}(S(Y)) := \mathrm{E}\left[\hat{S}(Y) - S^0\right] = \mathrm{E}\left[\hat{S}(Y)\right] - S^0$$

and the covariance matrix is given as

$$\mathrm{Var}(\hat{S}(Y)) = \mathrm{E}\left[\left(\hat{S}(Y) - \mathrm{E}\left[\hat{S}(Y)\right]\right)\left(\hat{S}(Y) - \mathrm{E}\left[\hat{S}(Y)\right]\right)^T\right].$$

Proof. The proof of the decomposition is an exercise I would like to execute at this point.

$$\begin{aligned}
\mathrm{MSE}(\hat{S}(Y)) &:= \mathrm{E}\left[\|\hat{S}(Y) - S^0\|^2\right] \\
&= \mathrm{trace}\left(\mathrm{E}\left[\left(\hat{S}(Y) - S^0\right)\left(\hat{S}(Y) - S^0\right)^T\right]\right) \\
&= \mathrm{trace}\left(\mathrm{E}\left[\hat{S}(Y)\hat{S}(Y)^T\right] - 2\mathrm{E}\left[\hat{S}(Y)\right] {S^0}^T + S^0 {S^0}^T\right) \\
&= \mathrm{trace}\left(\mathrm{E}\left[\hat{S}(Y)\hat{S}(Y)^T\right] - \mathrm{E}\left[\hat{S}(Y)\right]\mathrm{E}^T\left[\hat{S}(Y)\right]\right. \\
&\quad \left. + \mathrm{E}\left[\hat{S}(Y)\right]\mathrm{E}^T\left[\hat{S}(Y)\right] - 2\mathrm{E}\left[\hat{S}(Y)\right] {S^0}^T + S^0 {S^0}^T\right) \\
&= \mathrm{trace}\left(\mathrm{Var}\left(\hat{S}(Y)\right)\right) + \mathrm{trace}\left(\mathrm{Bias}(\hat{S}(Y))\mathrm{Bias}^T(\hat{S}(Y))\right) \\
&= \mathrm{trace}\left(\mathrm{Var}\left(\hat{S}(Y)\right)\right) + \left\|\mathrm{Bias}(\hat{S}(Y))\right\|^2
\end{aligned}$$

□

The *bias* variance decomposition gives an explanation for the so called *bias variance dilemma* in estimation. Both are quantities one would like to keep as low as possible. Whenever the *bias* declines the variance grows and vice versa under the assumption of a fixed *MSE*.
Since the quantile estimator for linear mixed models is already proven to be consistent – see Theorem 3.1 –, I would like to discuss the implication of that property on the *MSE* first. Consistency is a term, which is important in asymptotics. Thus there will be only an implication on the *MSE*, whenever the sample size n goes to infinity. For the relationship between consistency and *MSE* see the following theorems. First it can be shown that convergence of the *MSE* to zero implies the consistency.

Theorem 3.5. *The convergence of the MSE of an estimator to zero implies the consistency.*

Proof. The proof is an exercise and can be fulfilled as follows.
Let us assume for an estimator $\hat{S}_n(Y)$ and the true value S^0 that

$$\mathrm{MSE}(\hat{S}_n(Y)) \overset{n\to\infty}{\to} 0,$$

which implies by Theorem 3.3 that

$$\mathrm{Var}(\hat{S}_n(Y)) \overset{n\to\infty}{\to} 0. \tag{3.26}$$

Now let $\epsilon > 0$ be arbitrary and fixed. Then

$$\begin{aligned}
&P\left(\left|\hat{S}_n(Y) - S^0\right| \geqslant \epsilon\right)\\
&\quad = P\left(\left|\hat{S}_n(Y) - E\left[\hat{S}_n(Y)\right] + E\left[\hat{S}_n(Y)\right] - S^0\right| \geqslant \epsilon\right)\\
&\quad \overset{(\star)}{\leqslant} P\left(\left|\hat{S}_n(Y) - E\left[\hat{S}_n(Y)\right]\right| \geqslant \epsilon\right) P\left(\left|E\left[\hat{S}_n(Y)\right] - S^0\right| \geqslant \epsilon\right)\\
&\quad = P\left(\left|\hat{S}_n(Y) - E\left[\hat{S}_n(Y)\right]\right| \geqslant \epsilon\right)\\
&\quad \overset{(\star\star)}{\leqslant} \frac{\mathrm{Var}(\hat{S}_n(Y))}{\epsilon^2},
\end{aligned} \tag{3.27}$$

where $(\star)$ follows from the *Cauchy-Schwarz Inequality* and $(\star\star)$ follows from *Chebychev's Inequality* . By (3.26) the right hand side of (3.27) converges to zero as $n \to \infty$, which implies the consistency of $\hat{S}_n(Y)$. □

The other direction between consistency and *MSE* does not hold in general. Only by assuming a dominating random variable M I can state the following theorem.

Theorem 3.6. *For a consistent estimator* $\hat{S}_n(Y)$ *with* $\left|\hat{S}_n(Y)\right| \leqslant M \in L^2$ *it holds*

$$\mathrm{MSE}(\hat{S}_n(Y)) \overset{n\to\infty}{\to} 0.$$

Proof. The proof follows from the *Lebesgue Theorem* (cf. Theorem 3.12 in Aman and Escher [2008]). □

By application of Theorem 3.6 I can state the following corollary for the quantile estimator in linear mixed models.

Corollary 3.7. *For the quantile estimator as defined in* 3.14 *it holds*

$$\mathrm{MSE}(\hat{Q}_{Y|x}(\tau)) \overset{n\to\infty}{\to} 0.$$

Thus the consistency of the quantile estimator in linear mixed models implies the convergence of the *mean squared error* – and thus the *bias* and variance – to zero as $D \to \infty$ and $n_i \to \infty$. The *MSE* of the quantile estimator for finite D and n_i cannot be calculated easily due to the non-analytic calculation of $\hat{\theta}$ in the first step of the quantile estimation. Note that the *maximum likelihood* estimator $\hat{\theta}$ and thus the predictor for the random effect $\hat{V}$ are biased. So the quantile estimator (2.8) is not unbiased. They are only asymptotically unbiased because of their consistency – see Theorem 3.6.

3.6 APPLICATIONS AND EXTENSIONS OF THE LINEAR QUANTILE MIXED MODEL

There are further applications of quantile regression in linear mixed models than the estimation of the conditional quantile. The first I am going to discuss in the quantile estimation for count data. There the data must be transformed in order to fit the model. As in the linear model this is similar fulfilled by *jittering* and a log-transformation of the data. In the result the quantile estimator for count data will keep the property of being consistent, which I show in the following Section.
After that I introduce a method called *Microsimulation via Quantiles*, a new approach for the estimation of properties, which are beyond the mean. So it is possible to estimate not only individual conditional quantiles (on x_{ij}) but also area quantiles. For example estimators of lower and upper quartiles of income in areas can be obtained.

3.6.1 *Linear Quantile Mixed Models for Count Data*

As already discussed in Section 2.7.2 the quantiles of count data must be integers due to the fact that counts themselves are integers. Since the linear quantile mixed model (3.12) is a model for continuous data, it is not directly applicable on counts. The count mean mixed model or *Poisson* mixed model for a discrete random variable is Y_{ij} given x_{ij} is given as

$$\exp(x_{ij}^T\beta + V_i), \quad i = 1, 2, \ldots, D; j = 1, 2, \ldots, n_i \tag{3.28}$$

with

$$V_i \overset{iid}{\sim} N(0, \sigma_V^2).$$

This mean model needs to be improved in order to estimate quantiles of Y_{ij} given x_{ij} for a fixed $\tau \in (0, 1)$, $Q_{Y_{ij}|x_{ij}}(\tau)$. This will be fulfilled by *jittering* the data as discussed in the following Section 3.6.1.1. The main idea is the same as for count data in linear models, where Machado and Santos Silva [2005] already showed the consistency of the quantiles of counts – see Section 2.7.2 for details. Here, the consistency of quantile estimators in linear mixed models, as proved in Theorem 3.1, implies the consistency of the quantiles of counts. This is stated and demonstrated in Theorem 3.11.

3.6.1.1 *Jittering the Count Data*

The observations Y_{ij} $(i = 1, 2, \ldots, D; j = 1, 2, \ldots, n)$ are discrete and in linear models Machado and Santos Silva [2005] had the idea of *jittering* in order to get continuous data as described in Section 2.7.2.1. This method also works in the linear mixed model. By adding a standard *uniform* random variable U_{ij} independent from Y_{ij}, x_{ij}, and V_i we get a continuous observation Z_{ij}:

$$Z_{ij} := Y_{ij} + U_{ij}. \tag{3.29}$$

On this continuous random variable Z_{ij} we can apply the linear quantile mixed model (3.12). The quantile of the jittered data Z_{ij} is stated in the following theorem.

Theorem 3.8. *For a fixed $\tau \in (0,1)$ the quantile of Z_{ij} as defined in (3.29) is said to be*

$$Q_{Z_{ij}|x_{ij}}(\tau) = \exp(x_{ij}^T\beta + V_i) + \tau.$$

Proof. Let $\tau \in (0,1)$ be fixed. For a continuous random variable $Y_{ij} + U(-\tau, 1-\tau)$, where the mean model (3.28) holds for Y_{ij} the τ-quantile is

$$\begin{aligned} & Q_{Y_{ij}+U(-\tau,1-\tau)|x_{ij}}(\tau) = \exp(x_{ij}^T\beta + V_i) \\ \iff \quad & Q_{Y_{ij}+U(-\tau,1-\tau)+\tau|x_{ij}}(\tau) = \exp(x_{ij}^T\beta + V_i) + \tau \\ \iff \quad & Q_{Y_{ij}+U(0,1)|x_{ij}}(\tau) = \exp(x_{ij}^T\beta + V_i) + \tau. \end{aligned}$$

□

3.6.1.2 *Transformation of the Jittered Data*

In order to be able to apply the quantile estimation approach of linear quantile mixed models 3.12 there is need to transform the jittered data Z_{ij}. This is similar to the approach in the linear model in Section 2.7.2.2 and is for a fixed $\tau \in (0,1)$ fulfilled as follows

$$T(Z_{ij}, \tau) := \begin{cases} \log(\zeta), Z_{ij} \leqslant \tau \\ \log(Z_{ij} - \tau), Z_{ij} > \tau \end{cases}$$

with a small value ζ. This transformation is almost a continuous function and $\log(\zeta)$ is just the function value for negative values for $Z_{ij} - \tau$, since the logarithm is not defined for negative values. Therefore it follows for the transformed jittered data

$$T^{-1}(Z_{ij}, \tau) \approx \exp(Z_{ij}) + \tau$$

and hence I can state the following corollary.

Corollary 3.9. *The quantile of the transformed jittered data is given as*

$$Q_{T(Z_{ij},\tau)|x_{ij}}(\tau) = x_{ij}^T\beta_\tau + V_i.$$

Proof. The transformation T is almost continuous and thus it holds that

$$Q_{T(Z_{ij},\tau)|x_{ij}}(\tau) = T\left(Q_{Z_{ij}|x_{ij}}(\tau)\right).$$

In Theorem 3.8 it was shown that

$$Q_{Z_{ij}|x_{ij}}(\tau) = \exp(x_{ij}^T\beta + V_i) + \tau,$$

which implies that

$$\begin{aligned} Q_{T(Z_{ij},\tau)|x_{ij}}(\tau) &= T\left(\exp(x_{ij}^T\beta + V_i) + \tau, \tau\right) \\ &= \exp(x_{ij}^T\beta + V_i). \end{aligned}$$

□

3.6.1.3 *Applying Quantile Estimation in the Linear Mixed Model on the Transformed Jittered Data*

The transformed jittered data

$$Y_{ij}^\star := T(Z_{ij}, \tau)$$

is now continuous and we can apply the quantile estimation in linear mixed models as introduced in Section 3.3. There we estimate β_τ and predict V. In order to average out the error, which is based on the *jittering*, we apply an *averaged jittering*. That means we jitter our data M times and repeat the estimation of β_τ and V in each step. For M I would recommend $M \geqslant 10$. On the one hand a greater number of repeats M improves the estimation but it also extends the computational calculation time on the other hand. In the end we take the averaged estimators

$$\hat\beta_\tau = \frac{1}{M}\sum_{m=1}^{M}\hat\beta_{\tau,m} \quad \text{and} \quad \hat V = \frac{1}{M}\sum_{m=1}^{M}\hat V_m.$$

This leads to the quantile estimator of $Y_{ij}^\star$ given x_{ij}

$$\hat Q_{Y_{ij}^\star|x_{ij}}(\tau) = x_{ij}^T\hat\beta_\tau + \hat V_i, \quad i = 1, 2, \dots, D; j = 1, 2, \dots, n_i. \tag{3.30}$$

3.6.1.4 *Back-Transformation and Count Quantile*

From the τ-quantile of $Y_{ij}^\star$ we can calculate the τ-quantile of the observed counts Y_{ij} by the following theorem.

Theorem 3.10. *For a fixed* $\tau \in (0, 1)$ *the estimator for the* τ*-quantile of the observed counts* Y_{ij} *given* x_{ij} *is given by*

$$\begin{aligned} \hat Q_{Y_{ij}|x_{ij}}(\tau) &= \lceil T^{-1}(\hat Q_{Z_{ij}|x_{ij}}(\tau) - 1\rceil \\ &= \lceil \exp(x_{ij}^T\hat\beta_\tau + \hat V_i) + \tau - 1\rceil \end{aligned}$$

for $i = 1, 2, \dots, D$ *and* $j = 1, 2, \dots, n_i$.

Proof. The transformation T is almost continuous and bijective and thus it holds that

$$\hat{Q}_{Z_{ij}|x_{ij}}(\tau) = T^{-1}\left(\hat{Q}_{Y^{\star}_{ij}|x_{ij}}(\tau)\right).$$

Because of $Y_{ij} = Z_{ij} + U_{ij}$ with $U_{ij} \sim U(0,1)$ it also holds that

$$Y_{ij} - 1 \leqslant Z_{ij} - 1 \leqslant Y_{ij}.$$

Since the quantile function is non-decreasing, this implies for the conditional quantiles

$$\hat{Q}_{Y_{ij}|x_{ij}}(\tau) - 1 \leqslant \hat{Q}_{Z_{ij}|x_{ij}}(\tau) - 1 \leqslant \hat{Q}_{Y_{ij}|x_{ij}}(\tau).$$

The result now follows, because $\hat{Q}_{Y_{ij}|x_{ij}}(\tau)$ is an integer. □

3.6.1.5 *Consistency of the Quantile Estimation of Counts in Linear Mixed Models*

As in the linear model I can now state the following theorem about the consistency of the quantile estimator for counts.

Theorem 3.11. *For a fixed $\tau \in (0,1)$ the estimator for the τ-quantile of the observed counts Y_{ij} given x_{ij},*

$$\hat{Q}_{Y_{ij}|x_{ij}}(\tau) \quad i = 1,2,\ldots,D; j = 1,2,\ldots,n_i,$$

as defined in Theorem 3.10, is consistent.

Proof. In Theorem 3.1 I already proved that a quantile estimator for continuous random variables as $Y^{\star}_{ij}$, $\hat{Q}_{Y^{\star}_{ij}|x_{ij}}(\tau)$ as in (3.30), is consistent. Thus for a fixed $\epsilon > 0$ it holds that

$$P\left(\left|\hat{Q}_{Y^{\star}_{ij}|x_{ij}}(\tau) - Q_{Y^{\star}_{ij}|x_{ij}}(\tau)\right| \geqslant \epsilon\right) \overset{n\to\infty}{\to} 0.$$

Now the back-transformation T^{-1} and the ceiling function are continuous functions and it follows with Theorem 3.10 the consistency of $\hat{Q}_{Y_{ij}|x_{ij}}(\tau)$. □

3.6.1.6 *The Conclusion of Quantiles for Counts*

In this part I showed that the idea of *jittering* count data also works in linear mixed models. Thus one is able to estimate quantiles of count data by applying the quantile estimation in linear mixed models described in Section 3.3. This method works on continuous data. That is why the count data needed to be made continuous by the *jittering* and a transformation in order to have a linear quantile mixed model as in (3.12). After the estimation a back-transformation of the quantile estimators of the transformed jittered data gives the quantiles of the

counts.
Furthermore in Theorem 3.11 I showed that the quantile estimation in count data is consistent. This is implied by the consistency of the quantile estimation in linear mixed models, which was proven in the main theorem of this Chapter – Theorem 3.1. However this property of the quantile estimator in counts is a new improvement.

3.6.2 *Microsimulation via Quantiles*

The idea of Microsimulation via Quantiles is new. There has been no other literature but this thesis on the topic besides in Weidenhammer et al. [2016].
In practice there are parameters of interest in one area or overall observations, which are beyond mean estimation. In the linear mixed model (3.1) the predictor of Y given x as given in (3.9) is a mean predictor for the j^{th} unit in area i. The area mean $\hat{\bar{Y}}_i$ can then be given as the averaged means

$$\hat{\bar{Y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{Y}_{ij}$$

or for the samples units ($j \in S_i$) and the non-sampled units ($j \in R_i$) in area i

$$\hat{\bar{Y}}_i = \frac{1}{N_i} \left(\sum_{j \in S_i} Y_{ij} + \sum_{j \in R_i} \hat{Y}_{ij} \right).$$

Thus the mean of an area is the mean of all mean predictors. Similar the overall mean can be given as

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i=1}^{D} \sum_{j=1}^{N_i} \hat{Y}_{ij}.$$

This is totally different in quantile estimation. Equation 3.14 gives the conditional τ-quantile for the j^{th} unit in area i, from which one cannot derive the τ-quantile of the whole area $\hat{Q}_{Y_i|x_i}(\tau)$ nor the overall τ-quantile $\hat{Q}_{Y|x}(\tau)$. The mean of quantiles is not the area quantile

$$\hat{Q}_{Y_i|x_i}(\tau) \neq \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{Q}_{Y_{ij}|x_{ij}}(\tau).$$

Nevertheless there is a way of estimating area quantiles and more parameters of interest, which is called *Microsimulation via Quantiles* (MvQ).

3.6.2.1 *The Idea of Microsimulation via Quantiles*

Between quantiles and the distribution of a random variable Y exists a natural relationship. The distribution function F_Y can be rewritten as

$$F_Y(y) = \min\{\tau | Q_Y(\tau) \geqslant y\}.$$

Thus the empirical distribution function can be rewritten as

$$\hat{F}_Y(y) = \min\left\{\tau | \hat{Q}_Y(\tau) \geqslant y\right\},$$

where $\hat{Q}_Y(\tau)$ are the empirical quantiles.
In linear mixed models the quantiles can be estimated as given in (3.14). This estimation is fulfilled on a fixed τ. Let me now estimate quantile estimators on a increasing grid of τ's $T_K := (\tau_1, \tau_2, \dots, \tau_K)^T$ with $\tau_k < \tau_{k+1}$ for all $k = 1, 2, \dots K$. This leads to an empirical distribution function of Y_{ij}, the outcome for the j^{th} unit in area i, as follows

$$\hat{F}_{Y_{ij}|x_{ij}}(y) = \min\left\{\tau_k | \hat{Q}_{Y_{ij}|x_{ij}}(\tau_k) \geqslant y, k = 1, 2, \dots, K\right\} \wedge 1, \quad (3.31)$$

which is also dependent on the choice of the grid T_K. Thus I am able to estimate the whole distribution of the j^{th} unit in area i by (3.31). This even gives me the distribution of Y within one area or the over all distribution, from which I am able to estimate every parameter of interest by a *Monte Carlo* simulation.
Note that in the described procedure the event of quantile crossing may occur. This may happen whenever we estimate the quantile in a regression for every τ_k separately. Thus the improvement of the method with regard to quantile crossing is an open topic for research.

3.6.2.2 *The Implementation of Microsimulation via Quantiles*

For a given grid of τ $T_K = (\tau_1, \tau_2, \dots, \tau_K)$, e.g. $T_{99} = (.01, .02, \dots, .99)$, I estimate the quantiles as described in Section 3.3. This gives us an $N \times K$-dimensional matrix

$$\begin{pmatrix} \hat{Q}_{Y_{11}|x_{11}}(\tau_1) & \hat{Q}_{Y_{11}|x_{11}}(\tau_2) & \dots & \hat{Q}_{Y_{11}|x_{11}}(\tau_K) \\ \hat{Q}_{Y_{12}|x_{12}}(\tau_1) & \hat{Q}_{Y_{12}|x_{12}}(\tau_2) & \dots & \hat{Q}_{Y_{12}|x_{12}}(\tau_K) \\ \vdots & \vdots & & \vdots \\ \hat{Q}_{Y_{DN_D}|x_{DN_D}}(\tau_1) & \hat{Q}_{Y_{DN_D}|x_{DN_D}}(\tau_2) & \dots & \hat{Q}_{Y_{DN_D}|x_{DN_D}}(\tau_K) \end{pmatrix}.$$

Each row of this matrix gives me an estimation of the distribution function of Y_{ij} given x_{ij} as given in (3.31). From each $\hat{F}_{Y_{ij}}$ I draw a *Monte Carlo* sample of size MC

$$\tilde{y}_{ij} = (\tilde{y}_{ij}^{(1)}, \tilde{y}_{ij}^{(2)}, \dots, \tilde{y}_{ij}^{(MC)})^T. \quad (3.32)$$

This represents a microsimulation of the outcome Y_{ij} of the j^{th} unit in area i. For the whole area i the *Monte Carlo* sample

$$\tilde{y}_i = (\tilde{y}_{i1}^{(1)}, \tilde{y}_{i1}^{(2)}, \dots, \tilde{y}_{i1}^{(MC)}, \dots, \tilde{y}_{iN_i}^{(1)}, \tilde{y}_{iN_i}^{(2)}, \dots, \tilde{y}_{iN_i}^{(MC)})^T.$$

is a microsimulation of size $N_i \cdot MC$. This sample is just the combination of all microsimulations given in (3.32) and gives an estimated distribution of the outcome of Y in area i. Similar to this approach I could draw a microsimulation of all units $\tilde{y}$ and areas by glueing the samples given in (3.32) for all j and i together.
From $\tilde{y}_i$ or $\tilde{y}$ I can estimate now every parameter of interest. This can be fulfilled by taking the empirical version of this parameter from $\tilde{y}_i$ or $\tilde{y}$. Say I want to know the area mean the estimator would be

$$\widehat{mean}_i = mean(\tilde{y}_i)$$

and the τ-quantile estimator in area i is

$$\hat{Q}_{Y_i|x_i}(\tau) = q_\tau(\tilde{y}_i),$$

where $q_\tau(\tilde{y}_i)$ is defined as the empirical τ-quantile of the vector $\tilde{y}_i$. In the same matter other parameters can be estimated from the microsimulated data $\tilde{y}$. This approach can also performed for linear models by setting the quantile estimators from (2.8) in the empirical distribution function $\hat{F}_{Y_i|x_i}$.

3.6.2.3 *Simulation Study of Microsimulation via Quantiles*

In a simulation study I produced from the linear mixed model 50 pseudo samples with $D = 500$ areas with sample sizes $n_i = 10$ individuals in each ($i = 1, 2, \dots, 500$). The model is

$$Y_{ij} = 2 + 0.8x_{ij} + V_i + \varepsilon_{ij},$$

where the independent variables x_{ij} come from a *uniform distribution* on $(0, 1)$

$$x_{ij} \sim U(0, 1), \quad i = 1, 2, \dots, 500; j = 1, 2, \dots, 10.$$

The random effect was drawn from a *normal distribution* with zero mean and variance $\sigma_V^2 = 0.3^2$

$$V_i \sim N(0, 0.3^2), \quad i = 1, 2, \dots, 500,$$

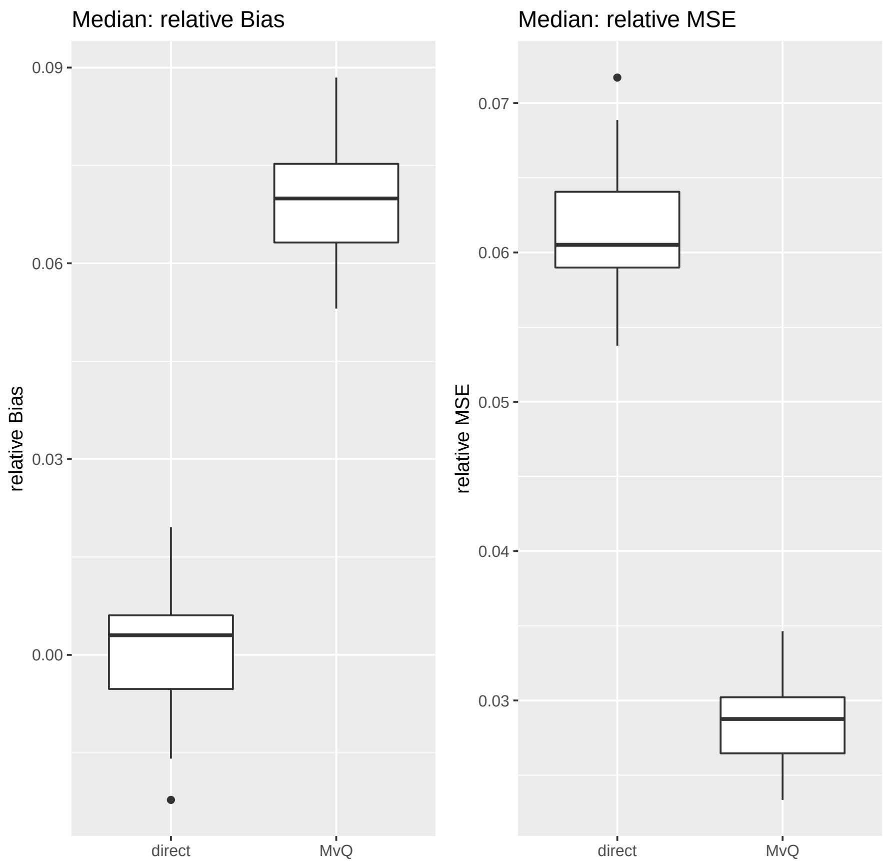

Figure 3.4: Boxplots of *relative MSE* and *relative bias* in median estimation by MvQ method of 50 simulation runs compared to the direct estimator

and the error term was drawn from a *transformed F distribution* with parameters $d_1 = 20$ and $d_1 = 20$

$$\varepsilon_{ij} \sim \sqrt{0.5 \frac{20 \cdot 18^2 \cdot 16}{2 \cdot 20^2 \cdot 38}} \left(F(20, 20) - \frac{20}{18} \right)$$

$$i = 1, 2, \ldots, 500; j = 1, 2, \ldots, 10$$

such that

$$E\left[\varepsilon_{ij}\right] = 0 \quad \text{and} \quad Var(\varepsilon_{ij}) = 0.5^2.$$

I chose the *transformed F distribution* in this simulation because it is heavy tailed. Therefore the estimation of upper quantiles is very fragile whenever the sample size is quite small.
The grid of τ I chose was $T_{99} = (0.01, 0.02, \ldots, 0.99)$. I estimated all quantiles for this grid with the package lqmm by Geraci [2016] in R. In the *Microsimulation via Quantiles* step I then estimated the area median and, as an upper quantile, the 95%-quantile in the area.

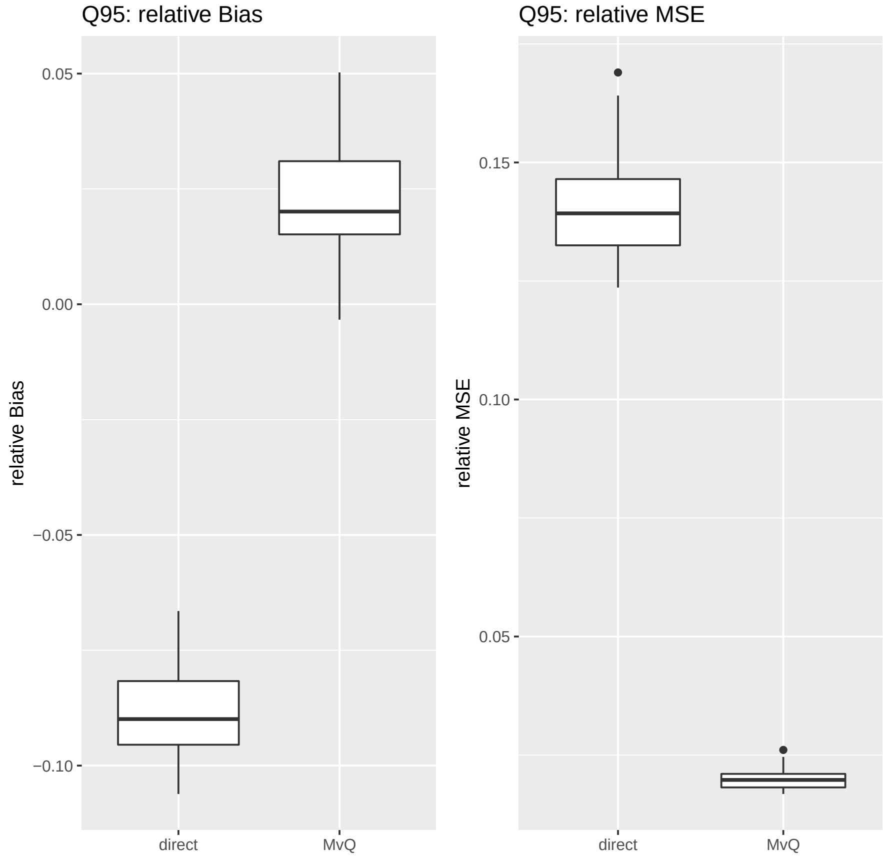

Figure 3.5: Boxplots of *relative MSE* and *relative bias* in 95%-quantile estimation by MvQ method of 50 simulation runs compared to the direct estimator

As measures of performance I calculated the *relative mean squared error* for each area $i = 1, 2, \dots, 500$

$$\mathrm{MSE}(\hat{m}_i) = \frac{1}{50} \sum_{\mathrm{Pop}=1}^{50} \frac{\left(\hat{m}_i^{\mathrm{Pop}} - m_i^{\mathrm{Pop}}\right)^2}{m_i^{\mathrm{Pop}}}$$

and the *relative bias*

$$\mathrm{Bias}(\hat{m}_i) = \frac{1}{50} \sum_{\mathrm{Pop}=1}^{50} \frac{\hat{m}_i^{\mathrm{Pop}} - m_i^{\mathrm{Pop}}}{m_i^{\mathrm{Pop}}},$$

where $\hat{m}_i^{\mathrm{Pop}}$ is the estimated value and m_i^{Pop} the true value (median or 95%-quantile) in the i^{th} area in Population $\mathrm{Pop} = 1, 2, \dots, 50$. As a reference estimator I chose the direct estimators of the median and the 95%-quantile, respectively. In Figures 3.4 and 3.5 boxplots of the *relative MSE* and the *relative bias* are drawn for the both estimates, respectively. In the median estimation the direct estimator is unbiased while the MvQ method has a small positive relative bias around 0.07.

This means the MvQ median estimator is about 7% higher than the real value. This behaviour may be healed by introducing a bias correction and is further discussed in Section 3.5. Nevertheless taking a look on the MSE the MvQ method outperforms the direct estimator. It has less than half the relative MSE. This means by applying the bias-variance decomposition that the MvQ method is more stable. This result is even better at the estimation of the 95%-quantile. As Figure 3.5 shows the relative MSE of the MvQ method is almost the same as for the median estimation while the direct estimator comes off worse. The direct estimator is not unbiased anymore. It is in average underestimated by about 8.5% while the MvQ estimator still has a positive bias of around 2%. This can be explained by the rather small sample size of $n = 10$ and the heavy tailed distribution of Y resulting from the *transformed F distribution* of the error term.
Altogether I can conclude that besides to the bias the MvQ method seems to be a stable estimation of quantiles. Especially the method performs better than the direct estimator whenever we want to estimate quantiles which are further on the edges. Due to the model assumption in the MvQ model we are even able to estimate quantiles for areas in which there is no observation which is not possible for the direct estimation.

3.6.2.4 *The Conclusion of Microsimulation via Quantiles*

In conclusion I can say that the method of *Microsimulation via Quantiles* provides first ideas for estimating parameter, which are beyond the mean. Thus it is possible to estimate area quantiles as I have simulated in the latter Section. Other parameters are possible. Since there the empirical distribution function is estimated, I get the distribution of the observation Y and may obtain any parameter of interest from that. This can be fulfilled by a *Monte Carlo* simulation. Then even the estimation of parameters like the Gini coefficient or poverty rates is possible.
Furthermore the MvQ method can be combined with the *jittering* introduced in Section 3.6.1. Hence parameters of interest of count data may also be estimated. Therefore the quantile estimators of the count data, which can be estimated as described before serve as the inverse of the empirical distribution function. From there everything else can be obtained by a *Monte Carlo* simulation.
On linear models with no random effect the method may also work by estimating the quantiles as described in Chapter 2. Then also other parameters than the mean can be estimated. Thus one is able to get quantile estimates unconditional on x_i. Remember the mean of a quantile is not a quantile. However *Microsimulation via Quantiles* is a way of estimating whatever one is interested in.

3.7 CONCLUSION

In this Chapter I introduced the idea of quantile estimation in linear mixed models. This approach is similar to the quantile regression in linear models, which was discussed in Chapter 2.
However there are differences in the estimation due to the existence of the random effect. As in the linear model the estimation of the regression parameter β_τ can be translated into a *maximum likelihood* estimation problem by assuming an *asymmetric Laplacian distributed* unit error model. There I was able to show the consistency of the estimation by using a theorem for *maximum likelihood* estimators with non-standard dependencies. The application of this *Weiss' Theorem* – see Theorem 4.1 – is such extensive that it is fulfilled in detail in the next Chapter 4.
The random effect in the linear mixed model also needs to be predicted, which is why the *maximum likelihood* estimation was just the first step in the quantile estimation. Hence the quantile estimator is a combination of the *maximum likelihood* estimator β_τ and the predictor of the random effect V. The latter prediction needs the *maximum likelihood* estimation in the first step. Eventually the consistency of that implies the consistency of the predictor, which in turn implies the consistency of the whole quantile estimation. This is proved and stated in the main theorem of this Chapter, Theorem 3.1.
In a simulation study in Section 3.4.2 the consistency became visible. There the *mean absolute deviation* between the quantile estimator and the true quantile decreased with growing within area sample sizes n_i and a growing number of areas.
Furthermore the consistency implied the asymptotic behaviour of the *mean squared error* of the quantile estimator. It converges to zero as the sample sizes and number of areas grow. However the quantile estimator is biased and only asymptotically unbiased.
This Chapter is ended with two applications of the quantile estimation in linear mixed model. In Section 3.6.1 I show that the consistency devolves to the quantile estimator for count data. In practice counts are quite common and thus there is need for their estimation. This can be implemented by the idea of *jittering*, which is described in detail.
The last extension of quantile estimation is the method *Microsimulation via Quantiles*, which I introduced in Section 3.6.2. This approach uses the interdependence between quantiles and the distribution function of a random variable. By the estimation of quantiles for a grid of $\tau \in (0, 1)$ the whole distribution can be estimated. From there any parameter of interest can be derived with an *Monte Carlo* approach. This is a microsimulation of the whole population, which explains the name of this method. First simulation results of this idea

are very promising, which is why the *Microsimulation via Quantiles* should be further examined in the future.

4 THE CONSISTENCY PROOF OF THE PARAMETER ESTIMATION

In this Chapter I am going to demonstrate the proof of the consistency of the parameter estimator, which is indispensable for the consistency of the quantile estimation. As a reminder, for a fixed τ given $\mathbf{X}$ the quantile estimator is given as

$$\hat{Q}_{Y|\mathbf{X}}(\tau) = \mathbf{X}\hat{\beta}_\tau + \mathbf{Z}\hat{V},$$

which follows from the linear quantile mixed model

$$Q_{Y|\mathbf{X}}(\tau) = \mathbf{X}\beta_\tau + \mathbf{Z}V.$$

This estimation of the quantile is conducted in two steps: First the parameters of the density of the n-dimensional observation vector Y are estimated by a *maximum likelihood* approach. As derived in Chapter 3 one may assume for the conditional distribution of Y given V an n-dimensional *asymmetric Laplace distribution*, which was discussed in Section 2.4.1

$$Y|V \sim ALD_n(\mathbf{X}\beta_\tau + \mathbf{Z}V, \sigma, \tau)$$

whilst in my case the distribution of the random effect vector is a D-dimensional *normal distribution*

$$V \sim N_D(0, \sigma_V \mathbf{I}_D).$$

This leads to a density function of Y as a function of the unknown parameter vector $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T$, which will be estimated in the first step as a *maximum likelihood* estimator of the log-likelihood of the vector Y.

In the second step a predictor for V is estimated using $\hat{\theta}$ from the *maximum likelihood* step, leading to an estimator of the quantile given by

$$\hat{Q}_{Y|\mathbf{X}}(\tau) = \mathbf{X}\hat{\beta}_\tau + \mathbf{Z}\hat{V}.$$

Thus also the consistency is shown in two steps. For the first step I need to show it for the *maximum likelihood* estimator $\hat{\theta}$. Since the n observations $Y_{1,1}, Y_{1,2}, \ldots, Y_{D,n_D}$ are not mutually independent of each other, I apply a theorem for this non-standard case. It is called *Weiss' Theorem* and will be introduced in Section 4.1.1 (cf. Weiss [1971] and Weiss [1973]). The application of this theorem is not straightforward. Hence its application – and thus the prove of the consistency of $\hat{\theta}$ – is divided into several steps. First I calculate the log-likelihood density of Y and its second derivatives, which is fulfilled in Section 4.1.3.

In the following Sections 4.2.1 and 4.2.2 the two assumptions of the *Weiss' Theorem* are shown for the linear quantile mixed model (3.12). This allows me to apply the theorem leading to the consistency of $\hat{\theta}$ under the assumptions (B1) to (B6) stated before Theorem 3.1. In the second step I need to show the consistency of the estimation of $\hat{V}$ from the consistent $\hat{\theta}$, which was already fulfilled in the proof of Theorem 3.1 leading to the consistency of the quantile estimator $\hat{Q}_{Y|\mathbf{X}}(\tau)$ for any given τ.

4.1 PRELIMINARIES

In the following I introduce the theorem, which I will apply in order to show the consistency in Section 4.1.1. Within this part I will also discuss the meaning of the assumptions made in the theorem in Section 4.1.1.2 and give examples how it was already applied in linear models with correlated error terms and in the linear mixed model with *normal* error terms in Section 4.1.1.3. Preliminaries for the application of the theorem follow in Section 4.1.3. In this part I am deriving the log-likelihood, which is needed for the theorem and their derivatives. After all tools have been provided the two main assumptions of the theorem are proven in Sections 4.2.1 and 4.2.2.

4.1.1 *The Weiss' Theorem*

The theorem by Lionel Weiss (1923-2000) is in common use, whenever the consistency of *maximum likelihood* estimators under non-standard cases are shown. It is a specialisation of the *Glivenko-Cantelli Theorem* (cf. Glivenko [1933] and Cantelli [1933]). The non-standard term applies here for the non-independence of the observations. Miller [1977] made first use of the theorem for *maximum likelihood* estimators in mixed models. Based on his work Pinheiro [1994] applied this theorem on the linear mixed model with *normal* errors in his doctoral thesis. In my case it will turn out to be more complicated than in the *normal* case but it follows mainly the same approach. The theorem was first introduced with three assumptions (see Weiss [1971]), which were later shortened to two assumptions (see Weiss [1973]). However Weiss [1973] is an addition to Weiss [1971], which is why both articles should be mentioned together.

4.1.1.1 *The Theorem*

Let me have a sample of random variables of size n $Y_1, Y_2, \ldots, Y_n$, which follow a distribution $P_{Y|\theta}$ dependent on an unknown k-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^k$. The sample vector is denoted by $Y = (Y_1, Y_2, \ldots, Y_n)^T$ whose Lebesgue density $f_Y(y|\theta)$ exists. Then the log-likelihood is defined by $\ell(\theta|y) = \log f_Y(y|\theta)$. Let

$\theta^0 \in \mathrm{int}(\Theta)$, where $\mathrm{int}(\Theta)$ is the interior of the k-dimensional set of unknown parameters Θ, be the true value for θ and there exist 2k sequences $K_1(n), K_2(n), \ldots, K_k(n)$, $M_1(n), M_2(n), \ldots, M_k(n)$ such that $K_\iota(n) \overset{n\to\infty}{\to} \infty$, $M_\iota(n) \overset{n\to\infty}{\to} \infty$, and $\frac{M_\iota(n)}{K_\iota(n)} \overset{n\to\infty}{\to} 0$ for all $\iota = 1, 2, \ldots, k$. The first assumption I state is

(A1)

$$-\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y)\bigg|_{\theta^0} \overset{P}{\to} B_{\iota_1,\iota_2}(\theta^0), \qquad (4.1)$$

where $\overset{P}{\to}$ means convergence in probability, whenever $n \to \infty$, $\iota_1, \iota_2 \in \{1, \ldots, k\}$, where $B_{\iota_1,\iota_2}(\theta^0)$ is continuous and the $k \times k$-dimensional matrix $B(\theta^0)$ is positive definite.

Before I state Assumption 2 I need to define some quantities. Let

$$\epsilon_{\iota_1,\iota_2}(\theta, \theta^0, n, Y) := -\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y) - B_{\iota_1,\iota_2}(\theta^0), \quad (4.2)$$

define the distance of $-\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y)$ to the limit stated in (A1) dependent on the sample vector Y, the sample size n, the parameter vector θ, and the true value θ^0. The set

$$N_n(\theta^0) := \left\{(\theta_1, \theta_2, \ldots, \theta_k)^T \,\middle|\, |\theta_\iota - \theta_\iota^0| \leqslant \frac{M_\iota(n)}{K_\iota(n)}, \iota \in \{1, 2, \ldots, k\}\right\} \quad (4.3)$$

consists of all parameters in Θ, which are within a range of $\frac{M(n)}{K(n)}$ around θ^0, and

$$R_n(\theta^0, \gamma) := \left\{Y \in \mathbb{R}^n \,\middle|\, \sum_{\iota_1=1}^{k}\sum_{\iota_2=1}^{k} M_{\iota_1}(n)M_{\iota_2}(n) \sup_{\theta\in N_n(\theta^0)} \left|\epsilon_{\iota_1,\iota_2}(\theta, \theta^0, n, Y)\right| \leqslant \gamma\right\} \quad (4.4)$$

is the set of values $Y \in \mathbb{R}^n$, for which the summed and by $M(n)$ weighted biggest distance defined in (4.2) for all $\theta \in N_n(\theta^0)$ is less than a given value γ. The second assumption can now be stated as follows:

(A2) Let there exist two positive non-random sequences $\{\gamma(n, \theta^0)\}$ and $\{\delta(n, \theta^0)\}$ with $\gamma(n, \theta^0) \overset{n\to\infty}{\to} 0$ and $\delta(n, \theta^0) \overset{n\to\infty}{\to} 0$ such that

$$P_\theta\left(R_n(\theta^0, \gamma(n, \theta^0))\right) > 1 - \delta(n, \theta^0) \quad \forall\theta \in N_n(\theta^0). \qquad (4.5)$$

Theorem 4.1 (The Weiss' Theorem (1971, 1973)). *For a sample $Y_1, Y_2, \ldots, Y_n$ as introduced before, where (A1) and (A2) holds it follows* *that there exists a sequence of maximum likelihood estimates* $\hat{\theta}(n)$*, which are roots of the equations* $\frac{\partial \ell(\theta|y)}{\partial \theta} = 0$ *such that*

$$\operatorname{diag}(K(n))(\hat{\theta}(n) - \theta^0) \xrightarrow{\mathcal{D}} N(0, B^{-1}(\theta^0)), \tag{4.6}$$

where $\xrightarrow{\mathcal{D}}$ *means the convergence in distribution as* $n \to \infty$ *and* $K(n) = (K_1(n), \ldots, K_k(n))^T$ *is the k-dimensional vector of the sequences* $K_\iota(n)$ *which were defined earlier.*

The *asymptotic normality* with rate $K(n)$, which is stated in this theorem implies the consistency of the maximum estimator $\hat{\theta}$, which will be shown in the end of this Chapter in Section 4.3. Furthermore it states the existence of the estimators themselves. However the existence gives no predication on their calculation.

4.1.1.2 *The Meaning of the Assumptions*

The sequence $K_\iota(n)$ is the convergence rate of the *asymptotic normality* (cf. Chapter 9.3 of van der Vaart [2007]). In the standard case with independently distributed observations it is $\sqrt{n}$ (cf. Aldrich [1997]). Thus in a non-standard case $K_\iota(n)$ cannot be any better, namely it can only be slower or of the same speed. There has to be a price for the dependencies. The left hand side of Assumption 1 as stated in (4.1)

$$-\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)} \frac{\partial^2}{\partial \theta_{\iota_1} \partial \theta_{\iota_2}} \ell(\theta|Y)\bigg|_{\theta^0} \tag{4.7}$$

turns out to be related to the negative mean value of the second derivative of the log-likelihood density, whenever $K_{\iota_1}(n)K_{\iota_2}(n) \to \infty$, which is fulfilled. In the independent case I could rewrite the log-likelihood as follows

$$\ell(\theta|y) = \sum_{i=1}^{n} \log\left(f_{Y_i}(y_i)\right).$$

Thus (4.7) is with $K_\iota(n) = \sqrt{n}$ the negative mean value of the second derivative of the log-likelihood density $\log\left(f_{Y_i}(y_i)\right)$. This is not possible in the case with dependent observations but the idea is the same. In the standard case I can apply the *Law of Large Numbers* and (4.7) converges in probability to the expected value

$$-E\left[\frac{\partial^2}{\partial \theta_{\iota_1} \partial \theta_{\iota_2}} \log\left(f_{Y_i}(y_i)\right)\bigg|_{\theta^0}\right], \tag{4.8}$$

which can be shown to be equal to the *Fisher Information* – see Lemma 4.2 – which is defined in Definition 2.5. In the standard case the inverse of the *Fisher Information* is the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}(n) - \theta^0)$ and it holds the equality to (4.8).

Lemma 4.2. *The Fisher Information matrix may be rewritten as*

$$I(\theta_0) = -E\left[\frac{\partial^2}{\partial\theta^2}\log\left(f_{Y_i}(y_i)\right)\bigg|_{\theta^0}\right].$$

The proof of this statement is an exercise from higher statistics classes. Nevertheless I am going to fulfil this calculation in the following in order to give the reader a deeper understanding.

Proof. The expected value of the matrix on the right hand side of the lemma can be further calculated as follows

$$\begin{aligned}
&E\left[\frac{\partial^2}{\partial\theta^2}\log\left(f_{Y_i}(Y_i)\right)\bigg|_{\theta^0}\right]\\
&= \int \frac{\partial^2}{\partial\theta^2}\log\left(f_{Y_i}(y_i)\right) f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&= \int \frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta} f_{Y_i}(y_i)}{f_{Y_{i}(y_i)}}\right) f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&= \int \left(\frac{\frac{\partial^2}{\partial\theta^2} f_{Y_i}(y_i)}{f_{Y_i}(y_i)} - \left(\frac{\frac{\partial}{\partial\theta} f_{Y_i}(y_i)}{f_{Y_i}(y_i)}\right)^2\right) f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&= \int \frac{\partial^2}{\partial\theta^2} f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i - \int \left(\frac{\partial}{\partial\theta}\log\left(f_{Y_i}(y_i)\right)\right)^2 f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&\overset{(\star)}{=} \frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}^{n_i}} f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&\qquad - \int \left(\frac{\partial}{\partial\theta}\log\left(f_{Y_i}(y_i)\right)\right)^2 f_{Y_i}(y_i)\bigg|_{\theta^0} dy_i\\
&\overset{(\star\star)}{=} \frac{\partial^2}{\partial\theta^2} 1 - E\left[\left(\frac{\partial}{\partial\theta}\log\left(f_{Y_i}(Y_i)\right)\right)^2\bigg|_{\theta^0}\right]\\
&= E\left[\left(\frac{\partial}{\partial\theta}\log\left(f_{Y_i}(Y_i)\right)\right)^2\bigg|_{\theta^0}\right] =: -I(\theta_0),
\end{aligned}$$

where $(\star)$ holds by application of the *Lebesgue Theorem* (cf. Theorem 3.12 in Aman and Escher [2008]) and $(\star\star)$ holds because f_{Y_i} is a probability density. □

In the non-standard case Assumption 2 (4.5) describes the speed of convergence in Assumption 1 (4.1). Within the proof of the *Weiss' Theorem* Assumption 2 is required for an intermediate step, where it is shown that

$$\frac{1}{K_\iota(n)}\frac{\partial}{\partial\theta_\iota}\ell(\theta|Y) \overset{\mathcal{D}}{\to} N(0, B(\theta^0)). \tag{4.9}$$

Assumption 1 is also needed within this step to show the convergence of the second derivatives to the covariance matrix $B(\theta^0)$, which serves as a version of the *Fisher Information* matrix in this case. By introducing the environment around θ^0 $N_n(\theta^0)$ in (4.3) Weiss is able to develop a *Taylor expansion* around θ^0 of the density of the observation Y $f_Y(Y|\theta)$. The set $R_n(\theta^0, \gamma)$ defined in (4.4) and the sequences $M_\iota(n)$ are needed for the convergence of the tail of the *Taylor expansion*. This step proves the existence of a relative maximum of $\ell(\theta, Y)$ in $N_n(\theta^0)$ named $\hat{\theta}(n)$. In the next step Weiss defines a sequence $\theta^\star(n)$ by

$$\text{diag}(K(n))(\theta^\star(n) - \theta^0) = \frac{1}{K_\iota(n)} \frac{\partial}{\partial \theta_\iota} \ell(\theta|Y)(B(\theta^0))^{-1},$$

which converges by (4.9) to $N\left(0, (B(\theta^0))^{-1}\right)$ in distribution. Finally he shows the equality of $\theta^\star(n)$ to the maximum $\hat{\theta}(n)$ by arguments of Cramér [1946]. Thus $\hat{\theta}(n)$ adopts the properties of $\theta^\star(n)$, especially

$$\text{diag}(K(n))(\hat{\theta}(n) - \theta^0) \xrightarrow{\mathcal{D}} N\left(0, (B(\theta^0))^{-1}\right).$$

The sequence $M_\iota(n)$ converges slower to ∞ than the convergence rate $K_\iota(n)$. Thus it holds that $M_\iota(n) \in o(K_\iota(n))$. It is only defined for constructive reasons.
Hence the two assumptions are needed for showing (4.9) and the existence of a maximum in a *Taylor expansion*. Alternatively one could also assume these two properties in order to show the statement of the *Weiss' Theorem* – see Theorem 4.1.

4.1.1.3 *Examples of Applications of the Weiss' Theorem*

The *Weiss' Theorem* was applied on a general linear model with dependencies within the error structure by Magnus [1978]. He showed the *asymptotic normality* of the regression coefficients and the variance parameters. In linear mixed models it was applied by Miller [1977] and Pinheiro [1994]. Their model has the assumption of *normal* random effects and *normal* individual error terms. They showed the *asymptotic normality* of the regression coefficient and the variance parameters in the random effect and the error term. Due to the *normal* assumptions, the log-likelihood in all mentioned cases can be calculated and therefore the two assumptions of the *Weiss' Theorem* were shown instantly with it. Since Miller [1977] and Pinheiro [1994] started with a linear mixed model, I will use the same structure of the proof in the following.

4.1.2 *Application of the Weiss' Theorem for Linear Quantile Mixed Models*

For a fixed $i = 1, 2, \dots, D$ the observations $Y_{i,1}, Y_{i,2}, \dots, Y_{i,n_i}$ are not mutually independent from each other due to the common random effect V_i. Another approach of consistency proofing is called for than in the general *maximum likelihood* approach (see Theorem 2.6). This

can be handled by applying the *Weiss' Theorem* – Theorem 4.1 – which is adapted on this non-standard case. As discussed in Section 4.1.1.2 the price for this case of non-independence can be found in the convergence rates of the consistency.
The application of the *Weiss' Theorem* in the linear quantile mixed model will be of the same structure as in Miller [1977] and Pinheiro [1994]. Compared to their approach the proof here is more complicated due to the *asymmetric Laplace distribution* for the individual error term. In the *normal* case one is able to use the fact that the convolution of two *normal distribution*s will lead into a *normal distribution*. Thus the density of Y as a sum of two *normal*s in this case is a *normal*. In contrast the density of Y in the linear quantile mixed model is a convolution of a *normal* and an *asymmetric Laplace distribution* , which is not analytically solvable. Thus it stays in the convolution form as an integral the whole time. Since the log-likelihood density is derived from this density and its second derivatives are needed, the *Lebesgue Theorem* is crucial to be able to interchange derivations and integration. This deviation is more complicated to compute. Moreover the results are not easily interpretable. Pinheiro [1994] could use results by Searle et al. [1992] for the second derivatives. Therefore he was able to skip this step.
However once the second derivatives are computed, I will stay with the structure of Miller [1977] and Pinheiro [1994], especially in the proof of Assumption 2 in Section 4.2.2. It will even turn out that the convergence rates are the same for the parameters σ_V and β_τ, which is a promising result for quantile regression in mixed models.
In principle the theorem can be applied as stated. An addition in the regression model is that the density f_Y is not only dependent on θ but also on the covariates in the design matrix $\mathbf{X}$ and therefore it should be named $f_{Y|\theta,\mathbf{X}}$. For notational simplicity I keep that in mind but drop $\mathbf{X}$ in the footnote. The unknown parameter vector is $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T$. Note that $\beta_\tau \in \mathbb{R}^p$ and thus $k = p + 2$. Nevertheless I will treat the parameter vector $\beta_\tau =: \theta_3$ as the third parameter I have to estimate. Hence the sequences $K_\iota(n)$ and $M_\iota(n)$ have only three instead of $k = p + 2$ different appearances each, where $K_3(n)$ will turn out to be the convergence rate of $\hat{\beta}_\tau(n)$. In the following the log-likelihood density and its derivatives are derived, such that I can proof the two assumptions.

4.1.3 *The Log-Likelihood Density and its Second Derivatives*

In the assumptions of the *Weiss' Theorem* I need the second derivatives of the log-likelihood density $\ell(\theta|Y)$ of Y. For that reason I am going to develop $\ell(\theta|Y)$ in Section 4.1.3.1. Next I calculate the first partial derivatives with respect to the parameters σ_V, σ, and β_τ in Section 4.1.3.2. The second partial derivatives have to be calculated with respect to all possible pairs from $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T$. This will lead to six different cases, which will be derived in Section 4.1.3.4. To maintain a coherent structure all results are stated in lemmata with the calculations in their proofs.

4.1.3.1 *The Log-Likelihood Density of* Y

The joint Lebesgue density of (Y, V) is the product of the conditional density function $f_{Y|V}$ and the density function of V

$$f_{Y,V}(y) = f_{Y|V}(y|\nu) \cdot f_V(\nu).$$

Thus the marginal density of Y can be obtained as the convolution by taking the integral with respect to the D-dimensional vector ν leading to the D-dimensional integral

$$\begin{aligned}
f_Y(y) &= \int_{\mathbb{R}^D} f_{Y,V}(y,\nu)d\nu \\
&= \int_{\mathbb{R}^D} \prod_{i=1}^{D} (f_{Y_i,V_i}(y_i,\nu_i))d\nu \\
&\overset{(\star)}{=} \int_{\mathbb{R}^D} \prod_{i=1}^{D} \left(\prod_{j=1}^{n_i} (f_{Y_{ij}|V_i}(y_{ij}|\nu_i)) f_{V_i}(\nu_i) \right) d\nu \\
&\overset{(\star\star)}{=} \prod_{i=1}^{D} \int_{\mathbb{R}} \prod_{j=1}^{n_i} (f_{Y_{ij}|V_i}(y_{ij}|\nu_i)) f_{V_i}(\nu_i) d\nu_i, \qquad (4.10)
\end{aligned}$$

where $(\star)$ holds because for a fixed i Y_{ij_1} given V_i is mutually independent from Y_{ij_2} given V_i for $j_1 \neq j_2$ and $(\star\star)$ holds by applying the *Theorem of Fubini* (cf. Klenke [2013], Satz 14.16).
The log-likelihood density for $\theta = (\sigma_V, \sigma, \beta_\tau^T)^T$ is then given by the logarithm of (4.10), hence

$$\begin{aligned}
\ell(\theta|Y) &= \log(f_Y(Y)) \\
&= \log\left(\prod_{i=1}^{D} \int_{\mathbb{R}} \prod_{j=1}^{n_i} (f_{Y_{ij}|V_i}(Y_{ij}|\nu_i)) f_{V_i}(\nu_i) d\nu_i \right) \\
&= \sum_{i=1}^{D} \log\left(\int_{\mathbb{R}} \prod_{j=1}^{n_i} (f_{Y_{ij}|V_i}(Y_{ij}|\nu_i)) f_{V_i}(\nu_i) d\nu_i \right). \qquad (4.11)
\end{aligned}$$

Equation (4.11) is the general form of the log-likelihood density in a linear mixed model. In the case of the linear mixed model with *normal* random effects and *normal* individual error terms, it will turn out to be a log-likelihood density of a *normal distribution*. In this case derivatives are relatively easy to calculate. In my case with *normal* random effects and individual error terms, which are *asymmetric Laplacian dis-*

tributed , it is more complicated because an analytical solution of the integral appears to be infeasible This is why the derivatives with respect to the unknown parameters are calculated directly from (4.11) in the following.

4.1.3.2 *The First Derivatives of the Log-Likelihood Density*

The derivatives of the log-likelihood density are calculated for each θ_ι from Θ with $\iota = 1, 2, 3$. Note that $\theta_3 = \beta_\tau$ is a p-dimensional vector and thus each derivative with respect to β_τ is a p-dimensional vector. Before obtaining all derivatives in these cases, I am giving a general form of it. This holds also for all linear mixed models because it is the derivative with respect to an unknown parameter θ_ι of the general log-likelihood density given in (4.11) is

$$
\begin{aligned}
\frac{\partial}{\partial\theta_\iota}\ell(\theta|Y) &= \frac{\partial}{\partial\theta_\iota}\sum_{i=1}^{D}\log\left(\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i\right)\\
&= \sum_{i=1}^{D}\frac{\frac{\partial}{\partial\theta_\iota}\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i}{\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i}\\
&\overset{(\star)}{=} \sum_{i=1}^{D}\frac{\int_{\mathbb{R}}\frac{\partial}{\partial\theta_\iota}\left(\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)\right)d\nu_i}{\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i}\\
&\overset{(\star\star)}{=} \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}}\frac{\partial}{\partial\theta_\iota}\left(\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))\right)f_{V_i}(\nu_i)d\nu_i}{\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i}\right.\\
&\qquad\left.+\frac{\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))\frac{\partial}{\partial\theta_\iota}f_{V_i}(\nu_i)d\nu_i}{\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i}\right), \quad (4.12)
\end{aligned}
$$

where $(\star)$ is an application of the *Lebesgue Theorem* and $(\star\star)$ holds by the *product rule*. Note that the denominator is the marginal density of Y_i:

$$\int_{\mathbb{R}}\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)d\nu_i = f_{Y_i}(Y_i).$$

The set of unknown model parameters, Θ, can be split into $\Theta_V := \{\sigma_V\}$ and $\Theta_Y := \{\sigma, \beta_\tau\}$ which define the density of V f_V and the density Y given V $f_{Y|V}$, respectively. Due to these impacts on the densities some derivatives are equal to zero and (4.12) can be simplified to

$$\frac{\partial}{\partial\theta_\iota}\ell(\theta|Y) = \sum_{i=1}^{D} \frac{\int_{\mathbb{R}} \frac{\partial}{\partial\theta_\iota}\left(\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))\right) f_{V_i}(\nu_i)d\nu_i}{f_{Y_i}(Y_i)} \quad \forall\theta_\iota \in \Theta_Y \tag{4.13}$$

and

$$\frac{\partial}{\partial\theta_\iota}\ell(\theta|Y) = \sum_{i=1}^{D} \frac{\int_{\mathbb{R}} \prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i)) \frac{\partial}{\partial\theta_\iota} f_{V_i}(\nu_i)d\nu_i}{f_{Y_i}(Y_i)} \quad \forall\theta_\iota \in \Theta_V \tag{4.14}$$

in the two cases. In the following the forms of each partial first derivative are derived.

4.1.3.2.1 Derivative with Respect to σ_V

Lemma 4.3. *The partial first derivative of (4.11) with respect to* σ_V *is given by*

$$\frac{\partial}{\partial\sigma_V}\ell(\theta|Y) = \sum_{i=1}^{D} E\left[\frac{V_i^2 - \sigma_V^2}{\sigma_V^3}\middle| Y_i\right]. \tag{4.15}$$

Proof. The parameter σ_V determines only the density f_V of the random effect and therefore has the form of (4.14). f_{V_i} is a density of a *normal distribution* with zero mean and variance equal to σ_V^2. Thus the first derivative of $f_{V_i}(\nu_i)$ with respect to σ_V is given by

$$\begin{aligned}
&\frac{\partial}{\partial\sigma_V} f_{V_i}(\nu_i) \\
&= \frac{\partial}{\partial\sigma_V}\left(\frac{1}{\sqrt{2\pi}\sigma_V}\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right)\right) \\
&= \frac{\partial}{\partial\sigma_V}\left(\frac{1}{\sqrt{2\pi}\sigma_V}\right)\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right) + \frac{1}{\sqrt{2\pi}\sigma_V}\frac{\partial}{\partial\sigma_V}\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right) \\
&= -\frac{1}{\sqrt{2\pi}\sigma_V^2}\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right) + \frac{1}{\sqrt{2\pi}\sigma_V}\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right)\frac{\nu_i^2}{\sigma_V^3} \\
&= \left(\frac{\nu_i^2}{\sigma_V^3} - \frac{1}{\sigma_V}\right)\frac{1}{\sqrt{2\pi}\sigma_V}\exp\left(-\frac{\nu_i^2}{2\sigma_V^2}\right) \\
&= \frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_V(\nu_i).
\end{aligned}$$

Using (4.14) the derivative of the likelihood is given by

$$\frac{\partial}{\partial \sigma_V}\ell(\theta|Y) = \sum_{i=1}^{D} \frac{\int_{\mathbb{R}} \prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3} f_{V_i}(\nu_i)d\nu_i}{f_{Y_i}(Y_i)}$$

$$= \sum_{i=1}^{D} \int_{\mathbb{R}} \frac{\nu_i^2-\sigma_V^2}{\sigma_V^3} \frac{\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)}{f_{Y_i}(Y_i)} d\nu_i$$

with the second fraction of densities simplified to

$$\frac{\prod_{j=1}^{n_i}(f_{Y_{ij}|V_i}(Y_{ij}|\nu_i))f_{V_i}(\nu_i)}{f_{Y_i}(Y_i)} = \frac{f_{Y_i,V_i}(Y_i,\nu_i)}{f_{Y_i}(Y_i)} = f_{V_i|Y_i}(\nu_i) \tag{4.16}$$

which is the density of the measure of V_i given Y_i. Hence I can write

$$\frac{\partial}{\partial \sigma_V}\ell(\theta|Y) = \sum_{i=1}^{D} \int_{\mathbb{R}} \frac{\nu_i^2-\sigma_V^2}{\sigma_V^3} f_{V_i|Y_i}(\nu_i)d\nu_i$$

$$= \sum_{i=1}^{D} E\left[\frac{V_i^2-\sigma_V^2}{\sigma_V^3}\middle| Y_i\right].$$

□

Note that (4.15) is also the form of the first derivative with respect to σ_V^2 in other distribution models of Y|V, whenever the distribution of the random effect is a *normal*. For example in a linear mixed model with *normal* individual error terms it has the same form.

Remark 4.4. *Since* V *is a vector and*

$$V^T V = \sum_{i=1}^{D} V_i^2,$$

I can also rewrite $\frac{\partial}{\partial \sigma_V}\ell(\theta|Y)$ *in vector form as follows*

$$\frac{\partial}{\partial \sigma_V}\ell(\theta|Y) = \frac{1}{\sigma_V^3} E\left[V^T V|Y\right] - \frac{D}{\sigma_V}.$$

The vector expression is just another way of expressing the derivative and simplifies calculations below. Nevertheless, for the further differentiation I keep the form with the sum as in Lemma 4.3. The interpretation of the conditional expectation with respect to Y will be given in Section 4.1.3.3.

4.1.3.2.2 Derivatives with Respect to σ and β_τ

For the derivatives with respect to σ and β_τ I start from equation (4.13). The derivative of the joint density $f_{Y_i|V_i}$ – which is described as a product of densities of *asymmetric Laplace distributions* – is given by

$$
\begin{aligned}
&f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad = \prod_{j=1}^{n_i} f_{Y_{ij}|V_i}(Y_{ij}|\nu_i)\\
&\quad = \prod_{j=1}^{n_i}\left(\frac{\tau(1-\tau)}{\sigma}\exp\left(-\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\right)\\
&\quad = \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}\exp\left(-\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\\
&\quad = \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}\exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right).
\end{aligned}
\tag{4.17}
$$

Derivative with Respect to σ

Lemma 4.5. *The partial first derivative of (4.11) with respect to σ is given by*

$$
\frac{\partial}{\partial\sigma}\ell(\theta|Y) = \sum_{i=1}^{D} E\left[-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle|Y_i\right].
\tag{4.18}
$$

Proof. For the proof I start from the derivative of (4.17) with respect to σ which is calculated as follows

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma}f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad = \frac{\partial}{\partial\sigma}\left(\frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}\right.\\
&\qquad\qquad \left.\exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right)\right)\\
&\quad = \frac{\partial}{\partial\sigma}\left(\frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}\right)\\
&\qquad \exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right)\\
&\qquad + \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}}\\
&\qquad \frac{\partial}{\partial\sigma}\exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{-n_i \tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i+1}} \\
&\quad \exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right) \\
&\quad + \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \\
&\quad \exp\left(\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right) \\
&\qquad \left(-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right) \\
&= \left(-\frac{n_i}{\sigma}-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)(Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i))\right) \\
&\quad f_{Y_i|V_i}(Y_i|\nu_i) \\
&= \left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right) f_{Y_i|V_i}(Y_i|\nu_i).
\end{aligned}
$$

Inserting this result in (4.13) proves the lemma:

$$
\begin{aligned}
&\frac{\partial}{\partial\sigma}\ell(\theta|Y) \\
&= \sum_{i=1}^{D}\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right) \\
&\qquad\qquad \cdot f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i)d\nu_i \\
&\overset{(\star)}{=} \sum_{i=1}^{D}\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right) \\
&\qquad\qquad \cdot f_{V_i|Y_i}(\nu_i)d\nu_i \\
&= \sum_{i=1}^{D} E\left[-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle| Y_i\right],
\end{aligned}
$$

where $(\star)$ follows from equation (4.16). □

Remark 4.6. *I can also rewrite* $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ *in vector form as follows*

$$
\frac{\partial}{\partial\sigma}\ell(\theta|Y) = -\frac{n}{\sigma}+\frac{1}{\sigma^2}E\left[\left(\mathbb{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbb{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right].
$$

The indicator function of a vector is defined as the vector of the indicator functions and $\mathbb{1}_n$ is an n-dimensional vector containing ones.

Derivative with Respect to β_τ

Lemma 4.7. *The partial first derivative of (4.11) with respect to* β_τ *is given by*

$$\frac{\partial}{\partial \beta_\tau} \ell(\theta|Y) = \sum_{i=1}^{D} E\left[-\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + V_i\}} - \tau) x_{ij} \middle| Y_i \right] \tag{4.19}$$

Proof. For applying (4.13) I need the derivative of the conditional density $f_{Y_i|V_i}$, which is given in (4.17). First, the derivative with respect to the whole p-dimensional vector β_τ receiving a p-dimensional derivative vector is derived by

$$\begin{aligned}
&\frac{\partial}{\partial \beta_\tau} f_{Y_i|V_i}(Y_i|\nu_i) \\
&= \frac{\partial}{\partial \beta_\tau} \left(\frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \right. \\
&\qquad \left. \exp\left(\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \right) \\
&= \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \\
&\quad \frac{\partial}{\partial \beta_\tau} \left(\exp\left(\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \right) \\
&= \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \exp\left(\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \\
&\quad \cdot \left(\frac{1}{\sigma} \sum_{j=1}^{n_i} \left(\frac{\partial}{\partial \beta_\tau} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right. \right. \\
&\qquad \left. \left. + (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau) \frac{\partial}{\partial \beta_\tau} (Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \right) \\
&= \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \exp\left(\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \\
&\quad \cdot \left(\frac{1}{\sigma} \sum_{j=1}^{n_i} \left(0 \cdot (Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) + (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(-x_{ij}) \right) \right) \\
&= \frac{\tau^{n_i}(1-\tau)^{n_i}}{\sigma^{n_i}} \exp\left(\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau)(Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)) \right) \\
&\quad \cdot \left(-\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau) x_{ij} \right)
\end{aligned}$$

$$= -\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}} - \tau)x_{ij} f_{Y_i|V_i}(Y_i|\nu_i).$$

By inserting this result in (4.13) I get

$$\begin{aligned}\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y) &= \sum_{i=1}^{D}\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}} - \tau)x_{ij}\right) \\ &\qquad \cdot f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i \\ &\overset{(\star)}{=} \sum_{i=1}^{D}\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}} - \tau)x_{ij}\right) \\ &\qquad \cdot f_{V_i|Y_i}(\nu_i)d\nu_i \\ &= \sum_{i=1}^{D} E\left[-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}} - \tau)x_{ij}|Y_i\right],\end{aligned}$$

where $(\star)$ follows from equation (4.16). □

Remark 4.8. *I can also rewrite* $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ *in vector form as follows*

$$\frac{\partial}{\partial\beta_{\tau,h}}\ell(\theta|Y) = -\frac{1}{\sigma}E\left[\mathbf{X}^T\left(\mathbb{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}} - \tau\mathbb{1}_n\right)|Y\right].$$

4.1.3.3 *The Conditional Expectation with Respect to* Y

All derivatives have in common that they can be depicted as conditional expectations with respect to the observation Y. These conditional expectation may be rewritten as integrals with respect to the measure $P_{V|Y}$. This is a somehow odd measure, in which I am normally not interested due to the non-observable vector of random effects, V. In my case of *asymmetric Laplacian distributed* error terms its density is not computable since the density of Y is not computable. In contrast to linear mixed models with *normal* error terms it is necessarily calculable. There it turns out to be a *normal distribution* with mean equal to $\sigma_V^2\mathbf{Z}^T\Sigma^{-1}(Y-\mathbf{X}\beta)$ and covariance matrix equal to $\sigma_V^2 I_D - \sigma_V^4\mathbf{Z}^T\Sigma^{-1}\mathbf{Z}$, where $\Sigma = \sigma_\varepsilon^2 I_n + \sigma_V^2\mathbf{Z}^T\mathbf{Z}$ is the covariance matrix of Y. Only the first derivative $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$ has no *asymmetric Laplace* *distribution* within its expression in (4.15). It has the same form as in the *normal* case and can be calculated there as follows

$$\begin{aligned}
\frac{\partial}{\partial \sigma_V}\ell(\theta|Y) &= \frac{1}{\sigma_V^3}\sum_{i=1}^{D} E[V_i^2|Y_i] - \frac{D}{\sigma_V} \\
&= \frac{1}{\sigma_V^3}\sum_{i=1}^{D}\left(Var(V_i|Y) + E^2[V_i|Y_i]\right) - \frac{D}{\sigma_V} \\
&\overset{(\star)}{=} \frac{1}{\sigma_V^3}trace(\sigma_V^2 I_D - \sigma_V^4 \mathbf{Z}^T\Sigma^{-1}\mathbf{Z}) \\
&\quad + \left(\sigma_V^2\mathbf{Z}^T\Sigma^{-1}\left(Y - \mathbf{X}\beta\right)\right)^T \sigma_V^2\mathbf{Z}^T\Sigma^{-1}\left(Y - \mathbf{X}\beta\right) - \frac{D}{\sigma_V} \\
&= -\sigma_V trace(\mathbf{Z}^T\Sigma^{-1}\mathbf{Z}) \\
&\quad + \sigma_V\left(\Sigma^{-1}(Y - \mathbf{X}\beta)\right)^T \mathbf{Z}\mathbf{Z}^T\Sigma^{-1}(Y - \mathbf{X}\beta),
\end{aligned}$$

where $(\star)$ follows from $P_{V|Y} = N(\sigma_V^2\mathbf{Z}^T\Sigma^{-1}\left(Y - \mathbf{X}\beta\right), \sigma_V^2 I_D - \sigma_V^4\mathbf{Z}^T\Sigma^{-1}\mathbf{Z})$. Compared to Searle et al. [1992] this is the same result for the first derivative. The other two derivatives have the form of the *asymmetric Laplace* case and therefore it is not possible to apply the conditional distribution of the *normal* case here.

All three derivatives have in common that they are conditional expectations. As a reminder these are random variables themselves and therefore must be treated accordingly. In application their realisations are set to zero in order to find the estimator $\hat{\theta}$:

$$\frac{\partial}{\partial \theta_\iota}\ell(\theta|Y) \overset{!}{=} 0, \quad \iota = 1, 2, 3.$$

Due to their appearance these equations have no analytical solution and are solved numerically, for example in the lqmm package for R by Geraci [2016].

4.1.3.4 *The Second Derivatives of the Log-Likelihood Density*

In this Section I am going to calculate the second derivatives of the log-likelihood density with respect to the unknown parameters θ_ι, $\iota = 1, 2, 3$. In the latter Section I calculated the three different first derivatives, where $\frac{\partial}{\partial \beta_\tau}\ell(\theta|Y)$ represents a p-dimensional vector of derivatives. Each of them must now be deviated with respect to all three different parameters σ_V, σ, and β_τ. This leads to nine different cases.

In the following part I show that the direction of deviation may be interchanged resulting in six different cases, which will be calculated below.

4.1.3.4.1 Interchange of Directions of Differentiation

By the *Schwarz' Theorem* (cf. Korollar 5.5 in Aman and Escher [2006]) I can interchange the directions of differentiation, whenever I deviate on an open set Θ with continuous second order derivatives. This can be satisfied, whenever I restrict Θ to

$$\Theta = (0, M)^2 \times (-M, M)^p$$

with $M > 0$ sufficiently large. This is not a strong assumption, because $\theta^0 \in \text{int}(\Theta)$ by (B2). Thus I can assume for the unknown parameter to fulfil $\theta^0 \in \Theta$. In the end of this Section I see that the second order derivatives are P-almost surely continuous. Thus the *Schwarz' Theorem* is applicable and it holds that

$$\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y) = \frac{\partial^2}{\partial\theta_{\iota_2}\partial\theta_{\iota_1}}\ell(\theta|Y) \quad \forall(\iota_1, \iota_2) \in \{1, 2, 3\}^2 . \tag{4.20}$$

Now I calculate all six cases with case numbers, which remain the same in later Sections. There I prove the validity of the two assumptions in the *Weiss' Theorem*. In each case I start with the general form of the derivative of the first derivative and give the expressions of the second derivatives for all six cases in lemmata.

4.1.3.4.2 The Derivatives of $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$

Generally the second derivative with respect to any parameter θ_ι of $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$ given in (4.15) is given by

$$\begin{aligned}
&\frac{\partial}{\partial\theta_\iota}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y) \\
&= \frac{\partial}{\partial\theta_\iota}\left(\sum_{i=1}^{D}\frac{\int_{\mathbb{R}}\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i}{f_{Y_i}(Y_i)}\right) \\
&\overset{(\star)}{=} \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}}\frac{\partial}{\partial\theta_\iota}\left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)\right)d\nu_i}{f_{Y_i}(Y_i)}\right. \\
&\qquad\left. - \frac{\int_{\mathbb{R}}\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i\frac{\partial}{\partial\theta_\iota}f_{Y_i}(Y_i)}{f_{Y_i}^2(Y_i)}\right),
\end{aligned} \tag{4.21}$$

where $(\star)$ is an application of the *Lebesgue Theorem*. The nominator in the first quotient of (4.21) can be rewritten as

$$\int_{\mathbb{R}} \frac{\partial}{\partial \theta_\iota} \left(\frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i) \right) d\nu_i$$
$$= \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta_\iota} \left(f_{Y_i|V_i}(Y_i|\nu_i) \right) \frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_{V_i}(\nu_i) \right.$$
$$\left. + f_{Y_i|V_i}(Y_i|\nu_i) \frac{\partial}{\partial \theta_\iota} \left(\frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_{V_i}(\nu_i) \right) \right) d\nu_i.$$

In this latter expression the first summand is zero, whenever $\theta_\iota \in \Theta_V$ and the second one is zero, whenever $\theta_\iota \in \Theta_Y$. The derivatives of $f_{Y_i}(Y_i)$ with respect to θ in the second quotient are indirectly given as the nominators in (4.15), (4.18), and (4.19) due to the equation

$$\frac{\partial}{\partial \theta_\iota} \ell(\theta|Y) = \sum_{i=1}^{D} \frac{\frac{\partial}{\partial \theta_\iota} f_{Y_i}(Y_i)}{f_{Y_i}(Y_i)}.$$

Thus I have for the derivatives $\frac{\partial}{\partial \theta_\iota} f_{Y_i}(Y_i)$ for $\theta_\iota \in \{\sigma_V, \sigma, \beta_\tau\}$

$$\frac{\partial}{\partial \sigma_V} f_{Y_i}(Y_i) = \int_{\mathbb{R}} \frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i) d\nu_i, \tag{4.22}$$

$$\frac{\partial}{\partial \sigma} f_{Y_i}(Y_i) = \int_{\mathbb{R}} \left(-\frac{n_i}{\sigma} + \frac{1}{\sigma} \sum_{j=1}^{n_i} \rho_\tau \left(\frac{Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)}{\sigma} \right) \right) f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i) d\nu_i, \tag{4.23}$$

and

$$\frac{\partial}{\partial \beta_\tau} f_{Y_i}(Y_i) = \int_{\mathbb{R}} \left(-\frac{1}{\sigma} \sum_{j=1}^{n_i} (\mathbf{1}_{\{Y_{ij} \leqslant x_{ij}^T \beta_\tau + \nu_i\}} - \tau) x_{ij} \right) f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i) d\nu_i, \tag{4.24}$$

respectively. These expressions are employed in all calculations of the second derivatives and will be referred to in the according cases.

Case 1: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma_V, \sigma_V)$

The first case is the second derivative of $\ell(\theta|Y)$ twice with respect to σ_V. Thus, I start with the first derivative $\frac{\partial}{\partial \sigma_V} \ell(\theta|Y)$ given in Lemma 4.3 and differentiate it with respect to σ_V leading to the following lemma.

Lemma 4.9. *The second derivative of* $\ell(\theta|Y)$ *with respect to* $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma_V, \sigma_V)$ *is given by*

$$\frac{\partial^2}{\partial \sigma_V^2}\ell(\theta|Y) = \frac{D}{\sigma_V^2} - \frac{3}{\sigma_V^4} E\left[V^T V|Y\right] + \frac{1}{\sigma_V^6} Var\left(V^T V|Y\right).$$

Proof. For $\theta_\iota = \sigma_V$ the second derivative $\frac{\partial}{\partial \theta_\iota}\left(\frac{\partial}{\partial \sigma_V}\ell(\theta|Y)\right)$ in (4.21) simplifies with the preliminary thoughts to

$$\frac{\partial^2}{\partial \sigma_V^2}\ell(\theta|Y) = \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}}\left(f_{Y_i|V_i}(Y_i|\nu_i)\frac{\partial}{\partial \sigma_V}\left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)\right)d\nu_i}{f_{Y_i}(Y_i)} - \frac{\int_{\mathbb{R}}\left(f_{Y_i|V_i}(Y_i|\nu_i)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)d\nu_i\,\frac{\partial}{\partial \sigma_V}f_{Y_i}(Y_i)}{f_{Y_i}^2(Y_i)}\right).$$

The derivative $\frac{\partial}{\partial \sigma_V}f_{Y_i}(Y_i)$ in the minuend was already stated in (4.22). So there is only need for the derivative in the subtrahend

$$\begin{aligned}
&\frac{\partial}{\partial \sigma_V}\left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)\\
&= \frac{\partial}{\partial \sigma_V}\left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\right)f_{V_i}(\nu_i) + \frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\frac{\partial}{\partial \sigma_V}f_{V_i}(\nu_i)\\
&= \frac{\partial}{\partial \sigma_V}\left(\frac{\nu_i^2}{\sigma_V^3}-\frac{1}{\sigma_V}\right)f_{V_i}(\nu_i) + \frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\\
&= \left(-\frac{3\nu_i^2}{\sigma_V^4}+\frac{1}{\sigma_V^2}\right)f_{V_i}(\nu_i) + \left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\right)^2 f_{V_i}(\nu_i)\\
&= \left(\frac{\sigma_V^2-3\nu_i^2}{\sigma_V^4} + \left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\right)^2\right)f_{V_i}(\nu_i).
\end{aligned}$$

Using these results it follows that

$$\frac{\partial^2}{\partial \sigma_V^2}\ell(\theta|Y) = \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}} f_{Y_i|V_i}(Y_i|\nu_i)\left(\frac{\sigma_V^2-3\nu_i^2}{\sigma_V^4}+\left(\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}\right)^2\right)f_V(\nu_i)d\nu_i}{f_{Y_i}(Y_i)} - \frac{\left(\int_{\mathbb{R}} f_{Y_i|V_i}(Y_i|\nu_i)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)d\nu_i\right)^2}{f_{Y_i}^2(Y_i)}\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{D} \left(\int_{\mathbb{R}} \frac{\sigma_V^2 - 3\nu_i^2}{\sigma_V^4} f_{V_i|Y_i}(\nu_i) d\nu_i \right. \\
&\quad + \int_{\mathbb{R}} \left(\frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} \right)^2 f_{V_i|Y_i}(\nu_i) d\nu_i \\
&\quad \left. - \left(\int_{\mathbb{R}} \frac{\nu_i^2 - \sigma_V^2}{\sigma_V^3} f_{V_i|Y_i}(\nu_i) d\nu_i \right)^2 \right) \\
&= \sum_{i=1}^{D} \left(E\left[\frac{\sigma_V^2 - 3V_i^2}{\sigma_V^4} \middle| Y_i \right] + E\left[\left(\frac{V_i^2 - \sigma_V^2}{\sigma_V^3} \right)^2 \middle| Y_i \right] \right. \\
&\quad \left. - E^2\left[\frac{V_i^2 - \sigma_V^2}{\sigma_V^3} \middle| Y_i \right] \right).
\end{aligned}
$$

In this expression each term of the sum can be further reduced in the following manner

$$
\begin{aligned}
&E\left[\frac{\sigma_V^2 - 3V_i^2}{\sigma_V^4} | Y_i \right] + E\left[\left(\frac{V_i^2 - \sigma_V^2}{\sigma_V^3} \right)^2 | Y_i \right] - E^2\left[\frac{V_i^2 - \sigma_V^2}{\sigma_V^3} | Y_i \right] \\
&= \frac{1}{\sigma_V^2} - \frac{3}{\sigma_V^4} E\left[V_i^2 | Y_i \right] \\
&\quad + \frac{1}{\sigma_V^6} \left(E\left[V_i^4 | Y_i \right] - 2\sigma_V^2 E\left[V_i^2 | Y_i \right] + \sigma_V^4 \right) \\
&\quad - \frac{1}{\sigma_V^6} \left(E^2\left[V_i^2 | Y_i \right] - 2\sigma_V^2 E\left[V_i^2 | Y_i \right] + \sigma_V^4 \right) \\
&= \frac{1}{\sigma_V^2} - \frac{3}{\sigma_V^4} E\left[V_i^2 | Y_i \right] + \frac{1}{\sigma_V^6} E\left[V_i^4 | Y_i \right] - \frac{1}{\sigma_V^6} E^2\left[V_i^2 | Y_i \right].
\end{aligned}
$$

The latter two summands can be rewritten as

$$
\begin{aligned}
&\sum_{i=1}^{D} \left[\frac{1}{\sigma_V^6} E\left[V_i^4 | Y_i \right] - \frac{1}{\sigma_V^6} E^2\left[V_i^2 | Y_i \right] \right] \\
&= \frac{1}{\sigma_V^6} \sum_{i=1}^{D} \left[E\left[V_i^4 | Y_i \right] - E^2\left[V_i^2 | Y_i \right] \right],
\end{aligned}
$$

where each summand is equal to the conditional variance of V_i^2 with respect to the measure $P_{V_i|Y_i}$, $Var\left(V_i^2|Y_i\right)$. Since the random effects V_i, $i = 1, 2, \ldots, D$, are pairwise independent, so are their squared transformations, leading to the covariances, $Cov(V_{i_1}^2, V_{i_2}^2)$, being equal to zero for all $i_1 \neq i_2$. Thus the sum of this variance is the variance of the sum

$$\frac{1}{\sigma_V^6}\sum_{i=1}^{D} Var\left(V_i^2|Y_i\right) = \frac{1}{\sigma_V^6} Var\left(\sum_{i=1}^{D} V_i^2 \middle| Y\right)$$
$$= \frac{1}{\sigma_V^6}\left(E\left[\left(\sum_{i=1}^{D} V_i^2\right)^2 \middle| Y\right] - E^2\left[\sum_{i=1}^{D} V_i^2 \middle| Y\right]\right).$$

Using the equation

$$V^T V = \sum_{i=1}^{D} V_i^2$$

I get as the result for a vector version of $\frac{\partial^2}{\partial \sigma_V^2}\ell(\theta|Y)$

$$\sum_{i=1}^{D}\left[\frac{1}{\sigma_V^2} - \frac{3}{\sigma_V^4}E\left[V_i^2|Y_i\right] + \frac{1}{\sigma_V^6}E\left[V_i^4|Y_i\right] - \frac{1}{\sigma_V^6}E^2\left[V_i^2|Y_i\right]\right]$$
$$\overset{(\star)}{=} \frac{D}{\sigma_V^2} - \frac{3}{\sigma_V^4}E\left[\sum_{i=1}^{D} V_i^2|Y\right] + \frac{1}{\sigma_V^6}E\left[\left(\sum_{i=1}^{D} V_i\right)^2 |Y_i\right]$$
$$- \frac{1}{\sigma_V^6}E^2\left[\sum_{i=1}^{D} V_i^2|Y\right]$$
$$= \frac{D}{\sigma_V^2} - \frac{3}{\sigma_V^4}E\left[V^T V|Y\right] + \frac{1}{\sigma_V^6}E\left[\left(V^T V\right)^2|Y\right] - \frac{1}{\sigma_V^6}E^2\left[V^T V|Y\right],$$

where $(\star)$ holds because of the pairwise independence of the random effects V_i, as discussed earlier. The latter terms may be rewritten as conditional expectation and variance, which leads to

$$\frac{\partial^2}{\partial \sigma_V^2}\ell(\theta|Y) = \frac{D}{\sigma_V^2} - \frac{3}{\sigma_V^4}E\left[V^T V|Y\right] + \frac{1}{\sigma_V^6}Var\left(V^T V|Y\right).$$

□

Thus the second derivative in case 1 is an expression of conditional expectations and variances with respect to the observation vector Y. Note that these expressions are random variables themselves. Further calculations of them will be executed in the proofing part in Section 4.2.

Case 2: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma_V, \sigma)$ or $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \sigma_V)$

The second case is now the derivative of $\frac{\partial}{\partial \sigma_V}\ell(\theta|Y)$ with respect to the second unknown parameter $\theta_2 = \sigma$. As shown in Section 4.1.3.4.1 this is the same expression as if one would differentiate $\frac{\partial}{\partial \sigma}\ell(\theta|Y)$ from Lemma 4.5 with respect to σ_V.

Lemma 4.10. *The second derivative of* $\ell(\theta|Y)$ *with respect to* $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \sigma_V)$ *or* $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \sigma_V)$ *is given by*

$$\frac{\partial^2}{\partial\sigma\partial\sigma_V}\ell(\theta|Y) = \frac{\partial^2}{\partial\sigma_V\partial\sigma}\ell(\theta|Y)$$
$$= \frac{1}{\sigma^2\sigma_V^3}\mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T \varepsilon, V^T V|Y\right).$$

Proof. The derivative of $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$ with respect to σ simplifies with (4.21) to

$$\frac{\partial}{\partial\sigma}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y) = \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}}\left(\frac{\partial}{\partial\sigma}\left(f_{Y_i|V_i}(Y_i|\nu_i)\right)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)d\nu_i}{f_{Y_i}(Y_i)}\right.$$
$$\left.-\frac{\int_{\mathbb{R}}\left(f_{Y_i|V_i}(Y_i|\nu_i)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)d\nu_i\frac{\partial}{\partial\sigma}f_{Y_i}(Y_i)}{f_{Y_i}^2(Y_i)}\right).$$

The proof of the derivative $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ of Lemma 4.5 and the derivative of f_Y with respect to σ given in (4.23) imply that

$$\frac{\partial}{\partial\sigma}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$$
$$= \sum_{i=1}^{D}\left(\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\right.$$
$$\cdot\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i|Y_i}(\nu_i)d\nu_i$$
$$-\int_{\mathbb{R}}\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i|Y_i}(\nu_i)d\nu_i$$
$$\left.\cdot\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{V_i|Y_i}(\nu_i)d\nu_i\right)$$
$$= \sum_{i=1}^{D}\left(E\left[\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\right)\cdot\frac{V_i^2-\sigma_V^2}{\sigma_V^3}\middle|Y_i\right]\right.$$
$$-E\left[\frac{V_i^2-\sigma_V^2}{\sigma_V^3}\middle|Y_i\right]$$
$$\left.\cdot E\left[-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle|Y_i\right]\right).$$

With the same argumentation as in the proof of Lemma 4.9 the derivative may be rewritten in vector form as

$$\begin{aligned}
&-\frac{n}{\sigma\sigma_V^3}\mathsf{E}\left[V^TV|Y\right]+\frac{Dn}{\sigma\sigma_V}\\
&+\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,V^TV|Y\right]\\
&-\frac{1}{\sigma^2\sigma_V}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right]\\
&+\frac{n}{\sigma\sigma_V^3}\mathsf{E}\left[V^TV|Y\right]-\frac{Dn}{\sigma\sigma_V}\\
&-\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right]\mathsf{E}\left[V^TV|Y\right]\\
&+\frac{1}{\sigma^2\sigma_V}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right],
\end{aligned}$$

which is by deduction equal to

$$\begin{aligned}
&\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,V^TV|Y\right]\\
&\quad-\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right]\mathsf{E}\left[V^TV|Y\right],
\end{aligned}$$

where $\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}$ is an n-dimensional indicator vector conditional on the n-dimensional logical vector $\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}$ and $\mathbf{1}_n$ is an n-dimensional vector of ones. With the transformation of the linear quantile mixed model (3.12) to the inverse model which is the model equation solved for V

$$V=\mathbf{Z}^{-1}(Y-\mathbf{X}\beta_\tau-\varepsilon) \tag{4.25}$$

the latter expression develops to

$$\begin{aligned}
&\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}\mathbf{Z}^{-1}(Y-\mathbf{X}\beta_\tau-\varepsilon)\}}-\tau\mathbf{1}_n\right)^T\right.\\
&\qquad\qquad\left.\left(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}\mathbf{Z}^{-1}(Y-\mathbf{X}\beta_\tau-\varepsilon))\right)V^TV|Y\right]\\
&-\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}\mathbf{Z}^{-1}(Y-\mathbf{X}\beta_\tau-\varepsilon)\}}-\tau\mathbf{1}_n\right)^T\right.\\
&\qquad\qquad\left.\left(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}\mathbf{Z}^{-1}(Y-\mathbf{X}\beta_\tau-\varepsilon))\right)|Y\right]\\
&\qquad\mathsf{E}\left[V^TV|Y\right]\\
&=\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon V^TV|Y\right]\\
&\quad-\frac{1}{\sigma^2\sigma_V^3}\mathsf{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\mathsf{E}\left[V^TV|Y\right].
\end{aligned}$$

This expression of the difference of the conditional expectation of the product and the product of the conditional expectation can be rewritten as the conditional covariance

$$\frac{1}{\sigma^2\sigma_V^3}\operatorname{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon, V^TV|Y\right),$$

where $\operatorname{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon, V^TV|Y\right)$ is the conditional covariance between the random variables $\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon$ and V^TV. □

Thus the second derivative in case 2 is again a conditional expression – the conditional covariance – with respect to the observation vector Y. In Section 4.2 I will further investigate this term.

Case 3: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1},\theta_{\iota_2}) = (\sigma_V,\beta_\tau)$ or $(\theta_{\iota_1},\theta_{\iota_2}) = (\beta_\tau,\sigma_V)$

Consequently, the third case is the derivative of $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$ with respect to the unknown p-dimensional parameter vector $\theta_3 = \beta_\tau$. As a result this expression is again p-dimensional. With the argumentation of Section 4.1.3.4.1 the result is the same expression as if one would differentiate $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ from Lemma 4.7 with respect to σ_V.

Lemma 4.11. *The second derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1},\theta_{\iota_2}) = (\sigma_V,\beta_\tau)$ or $(\theta_{\iota_1},\theta_{\iota_2}) = (\beta_\tau,\sigma_V)$ is given by*

$$\begin{aligned}\frac{\partial^2}{\partial\beta_\tau\partial\sigma_V}\ell(\theta|Y) &= \frac{\partial^2}{\partial\sigma_V\partial\beta_\tau}\ell(\theta|Y)\\ &= -\frac{1}{\sigma\sigma_V^3}\operatorname{Cov}\left(V^TV,\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}.\end{aligned}$$

Proof. The derivative of $\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)$ with respect to the parameter β_τ simplifies with (4.21) to

$$\begin{aligned}&\frac{\partial}{\partial\beta_\tau}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)\\ &= \sum_{i=1}^{D}\left(\frac{\int_{\mathbb{R}}\left(\frac{\partial}{\partial\beta_\tau}\left(f_{Y_i|V_i}(Y_i|\nu_i)\right)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)d\nu_i}{f_{Y_i}(Y_i)}\right.\\ &\qquad\left.-\frac{\int_{\mathbb{R}}\left(f_{Y_i|V_i}(Y_i|\nu_i)\frac{\nu_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i}(\nu_i)\right)d\nu_i\,\frac{\partial}{\partial\beta_\tau}f_{Y_i}(Y_i)}{f^2_{Y_i}(Y_i)}\right).\end{aligned}$$

With the proof of the derivative $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ of Lemma 4.7 and the derivative of f_Y with respect to β_τ given in (4.24) it follows that

$$\begin{aligned}
&\frac{\partial}{\partial\beta_\tau}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y)\\
&=\sum_{i=1}^{D}\left(\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+v_i\}}-\tau)x_{ij}\right)\frac{v_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i|Y_i}(v_i)dv_i\right.\\
&\quad-\int_{\mathbb{R}}\frac{v_i^2-\sigma_V^2}{\sigma_V^3}f_{V_i|Y_i}(v_i)dv_i\\
&\quad\left.\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+v_i\}}-\tau)x_{ij}\right)f_{V_i|Y_i}(v_i)dv_i\right)\\
&=\sum_{i=1}^{D}\left(E\left[\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}}-\tau)x_{ij}\right)\cdot\frac{V_i^2-\sigma_V^2}{\sigma_V^3}\middle|Y_i\right]\right.\\
&\quad\left.-E\left[\frac{V_i^2-\sigma_V^2}{\sigma_V^3}\middle|Y_i\right]E\left[-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}}-\tau)x_{ij}\middle|Y_i\right]\right).
\end{aligned}$$

Bringing this expression in vector form I get

$$\begin{aligned}
&-\frac{1}{\sigma\sigma_V^3}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}V^TV|Y\right]\\
&+\frac{1}{\sigma\sigma_V}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}|Y\right]\\
&+\frac{1}{\sigma\sigma_V^3}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}|Y\right]E\left[V^TV|Y\right]\\
&-\frac{1}{\sigma\sigma_V}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}|Y\right],
\end{aligned}$$

which is equal to

$$\begin{aligned}
&-\frac{1}{\sigma\sigma_V^3}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}V^TV|Y\right]\\
&\quad+\frac{1}{\sigma\sigma_V^3}E\left[(\mathbf{1}_{\{Y\leqslant\mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n)^T\mathbf{X}|Y\right]E\left[V^TV|Y\right].
\end{aligned}$$

By using the model (4.25) it follows that

$$\begin{aligned}
&-\frac{1}{\sigma\sigma_V^3}\left(E\left[(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n)^T\mathbf{X}V^TV|Y\right]\right.\\
&\qquad\left.-E\left[(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n)^T\mathbf{X}|Y\right]E\left[V^TV|Y\right]\right)\\
&\quad=-\frac{1}{\sigma\sigma_V^3}\left(E\left[V^TV(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n)^T|Y\right]\right.\\
&\qquad\qquad\left.-E\left[V^TV|Y\right]E\left[(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n)^T|Y\right]\right)\mathbf{X},
\end{aligned}$$

which can be expressed as a conditional covariance

$$
\begin{aligned}
&-\frac{1}{\sigma\sigma_V^3}\operatorname{Cov}\left(V^T V, (\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n)^T | Y\right)\mathbf{X} \\
&= -\frac{1}{\sigma\sigma_V^3}\operatorname{Cov}\left(V^T V, \mathbf{1}_{\{\varepsilon\leqslant 0\}}^T | Y\right)\mathbf{X},
\end{aligned}
$$

where $\operatorname{Cov}\left(V^T V, \mathbf{1}_{\{\varepsilon\leqslant 0\}}^T | Y\right)$ is a $1 \times n$ conditional covariance vector of the variable $V^T V$ and the vector $\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T$. □

4.1.3.5 *The Derivatives of* $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$

Generally, a derivative of $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ given in (4.18) with respect to a parameter θ_ι is given by

$$
\begin{aligned}
&\frac{\partial}{\partial\theta_\iota}\frac{\partial}{\partial\sigma}\ell(\theta|Y) \\
&= \frac{\partial}{\partial\theta_\iota}\left(\sum_{i=1}^{D}\int_{\mathbb{R}}\frac{1}{f_{Y_i}(Y_i)}\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right)\right. \\
&\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i)\, d\nu_i\right) \\
&\overset{(\star)}{=} \sum_{i=1}^{D}\left(\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\frac{\partial}{\partial\theta_\iota}\left(\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right)\right.\right. \\
&\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i)\right) d\nu_i \\
&\qquad - \frac{1}{f_{Y_i}^2(Y_i)}\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right) \\
&\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i)\, d\nu_i \frac{\partial}{\partial\theta_\iota} f_{Y_i}(Y_i)\right),
\end{aligned}
\tag{4.26}
$$

where $(\star)$ holds by applying the *Lebesgue Theorem*. The first integral of (4.26) can be rewritten as

$$
\begin{aligned}
&\int_{\mathbb{R}}\frac{\partial}{\partial\theta_\iota}\left(\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right)\right. \\
&\qquad\left. \cdot f_{Y_i|V_i}(Y_i|\nu_i) f_{V_i}(\nu_i)\right) d\nu_i
\end{aligned}
$$

$$= \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta_\iota} \left(\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma} \sum_{j=1}^{n_i} \rho_\tau \left(\frac{Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)}{\sigma} \right) \right) f_{Y_i|V_i}(Y_i|\nu_i) \right) \right.$$

$$\cdot f_{V_i}(\nu_i)$$

$$+ \left(-\frac{n_i}{\sigma} + \frac{1}{\sigma} \sum_{j=1}^{n_i} \rho_\tau \left(\frac{Y_{ij} - (x_{ij}^T \beta_\tau + \nu_i)}{\sigma} \right) \right) f_{Y_i|V_i}(Y_i|\nu_i)$$

$$\left. \cdot \frac{\partial}{\partial \theta_\iota} f_{V_i}(\nu_i) \right) d\nu_i.$$

In this expression the first summand is zero, whenever $\theta_\iota \in \Theta_V$ and the second one is zero, whenever $\theta_\iota \in \Theta_Y$. The derivatives of $f_{Y_i}(Y_i)$ in the second quotient are given in (4.22), (4.23), and (4.24). By application of the *Schwarz' Theorem* – see (4.20) – I have

$$\frac{\partial}{\partial \sigma_V} \frac{\partial}{\partial \sigma} \ell(\theta|Y) = \frac{\partial}{\partial \sigma} \frac{\partial}{\partial \sigma_V} \ell(\theta|Y) \tag{4.27}$$

which is the second derivative from case 2 and given in Lemma 4.10. Hence I only need to calculate with the derivative of $\frac{\partial}{\partial \sigma} \ell(\theta|Y)$ with respect to σ_V and β_τ, which will be cases 4 and 5.

Case 4: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \sigma)$

The fourth case is the second derivative of $\ell(\theta|Y)$ twice with respect to σ. I start with the first derivative $\frac{\partial}{\partial \sigma} \ell(\theta|Y)$ given in Lemma 4.5 and differentiate it with respect to σ leading to the following lemma.

Lemma 4.12. *The second derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \sigma)$ is given by*

$$\frac{\partial^2}{\partial \sigma^2} \ell(\theta|Y)$$
$$= \frac{n}{\sigma^2} - \frac{2}{\sigma^3} \left(E\left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T \varepsilon | Y \right] - \tau \mathbf{1}_n^T E\left[\varepsilon | Y\right] \right) + \frac{1}{\sigma^4} Var\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T \varepsilon | Y \right).$$

Proof. For the parameter σ the derivative of $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ simplifies with (4.26) and previous argumentation to

$$\begin{aligned}
&\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)\\
&= \sum_{i=1}^{D}\left(\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\frac{\partial}{\partial\sigma}\left(\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\right.\right.\\
&\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i)\right) f_{V_i}(\nu_i)d\nu_i\\
&\quad -\frac{1}{f_{Y_i}^2(Y_i)}\int_{\mathbb{R}}\left(\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\right.\\
&\qquad\left.\left. f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)\right)d\nu_i\frac{\partial}{\partial\sigma}f_{Y_i}(Y_i)\right).
\end{aligned}$$

The derivative in the first integral can be calculated using $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ given in (4.18)

$$\begin{aligned}
&\frac{\partial}{\partial\sigma}\left(\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i)\right)\\
&= \frac{\partial}{\partial\sigma}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\frac{\partial}{\partial\sigma}f_{Y_i|V_i}(Y_i|\nu_i)\\
&= \left(\frac{n_i}{\sigma^2}-\frac{2}{\sigma^2}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)^2 f_{Y_i|V_i}(Y_i|\nu_i)\\
&= \left(\frac{n_i}{\sigma^2}-\frac{2}{\sigma^2}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right.\\
&\quad\left.+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)^2\right)f_{Y_i|V_i}(Y_i|\nu_i).
\end{aligned}$$

Inserting this result and using Equation (4.23) leads to

$$\begin{aligned}
&\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)\\
&=\sum_{i=1}^{D}\left(\left(\int_{\mathbb{R}}\left(\frac{n_i}{\sigma^2}-\frac{2}{\sigma^2}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{V_i|Y_i}(\nu_i)d\nu_i\right.\right.\\
&\quad+\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)^2 f_{V_i|Y_i}(\nu_i)d\nu_i\\
&\quad-\left(\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\right.\\
&\qquad\left.\left. f_{V_i|Y_i}(\nu_i)d\nu_i\right)^2\right)\\
&=\sum_{i=1}^{D}\left(E\left[\frac{n_i}{\sigma^2}-\frac{2}{\sigma^2}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle|Y_i\right]\right.\\
&\quad+E\left[\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\right)^2\middle|Y_i\right]\\
&\quad\left.-E^2\left[-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle|Y_i\right]\right).
\end{aligned}$$

In vector notation the derivative $\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)$ is given by

$$\begin{aligned}
&\frac{n}{\sigma^2}-\frac{2}{\sigma^3}E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))|Y\right]\\
&+\frac{1}{\sigma^4}E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\right.\\
&\qquad\left.\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))|Y\right]\\
&-\frac{1}{\sigma^4}E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))|Y\right]\\
&\qquad E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))|Y\right]
\end{aligned}$$

When I make use of model (4.25) the derivative develops to

$$\begin{aligned}
&\frac{n}{\sigma^2}-\frac{2}{\sigma^3}E\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\\
&+\frac{1}{\sigma^4}\left(E\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\right.\\
&\qquad\left.-E\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]E\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{n}{\sigma^2} - \frac{2}{\sigma^3}\left(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}_n^T E\left[\varepsilon|Y\right]\right) \\
&\quad + \frac{1}{\sigma^4}\Big(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - 2\tau\mathbf{1}_n^T\varepsilon E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] + \tau^2\mathbf{1}_n^T\varepsilon\mathbf{1}_n^T\varepsilon \\
&\qquad - E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] + 2\tau\mathbf{1}_n^T\varepsilon E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] \\
&\qquad - \tau^2\mathbf{1}_n^T\varepsilon\mathbf{1}_n^T\varepsilon\Big) \\
&= \frac{n}{\sigma^2} - \frac{2}{\sigma^3}\left(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}_n^T E\left[\varepsilon|Y\right]\right) \\
&\quad + \frac{1}{\sigma^4}\left(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right) \\
&= \frac{n}{\sigma^2} - \frac{2}{\sigma^3}\left(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}_n^T E\left[\varepsilon|Y\right]\right) + \frac{1}{\sigma^4}Var\left(\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right),
\end{aligned}
$$

where $Var\left(\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right)$ is the conditional variance of the variable $\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon$, which is the sum of all positive error variables ε_{ij}. □

Case 5: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \beta_\tau)$ or $(\theta_{\iota_1}, \theta_{\iota_2}) = (\beta_\tau, \sigma)$

In the second differentiation of $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ I now calculate the derivative with respect to the parameter vector β_τ. As for case 3, this fifth case will have a p-dimensional vector as a result and is given in the following lemma.

Lemma 4.13. *The second derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\sigma, \beta_\tau)$ or $(\theta_{\iota_1}, \theta_{\iota_2}) = (\beta_\tau, \sigma)$ is given by*

$$
\begin{aligned}
\frac{\partial^2}{\partial\beta_\tau\partial\sigma}\ell(\theta|Y) &= \frac{\partial^2}{\partial\sigma\partial\beta_\tau}\ell(\theta|Y) \\
&= -\frac{1}{\sigma^2}\left(E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right] - \tau\mathbf{1}_n^T\right)\mathbf{X} \\
&\quad - \frac{1}{\sigma^3}Cov\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T\varepsilon, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}.
\end{aligned}
$$

Proof. For the parameter β_τ the derivative of $\frac{\partial}{\partial\sigma}\ell(\theta|Y)$ simplifies with (4.26) and previous argumentation to

$$
\begin{aligned}
&\frac{\partial}{\partial\beta_\tau}\frac{\partial}{\partial\sigma}\ell(\theta|Y) \\
&= \sum_{i=1}^{D}\Bigg(\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\frac{\partial}{\partial\beta_\tau}\left(\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right)\right. \\
&\qquad\qquad \left. f_{Y_i|V_i}(Y_i|\nu_i)\right) f_{V_i}(\nu_i)d\nu_i \\
&\quad - \frac{1}{f^2_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma} + \frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij} - (x_{ij}^T\beta_\tau + \nu_i)}{\sigma}\right)\right) \\
&\qquad\qquad f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i\frac{\partial}{\partial\beta_\tau}f_{Y_i}(Y_i)\Bigg).
\end{aligned}
$$

Further I can calculate the derivative in the first integral by using Lemma 4.7

$$\frac{\partial}{\partial\beta_\tau}\left(\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i)\right)$$
$$=\frac{\partial}{\partial\beta_\tau}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i)$$
$$+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)\frac{\partial}{\partial\beta_\tau}f_{Y_i|V_i}(Y_i|\nu_i)$$
$$=\left(-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{Y_i|V_i}(Y_i|\nu_i)$$
$$+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right)$$
$$\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{Y_i|V_i}(Y_i|\nu_i)$$
$$=\left(-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right.$$
$$+\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\right)$$
$$\left.\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\right)f_{Y_i|V_i}(Y_i|\nu_i).$$

Inserting this result in the first equation of this proof it follows that

$$
\begin{aligned}
&\frac{\partial}{\partial \beta_\tau}\frac{\partial}{\partial \sigma}\ell(\theta|Y) \\
&= \sum_{i=1}^{D}\left(\int_{\mathbb{R}}\left(-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{V_i|Y_i}(\nu_i)d\nu_i\right. \\
&\qquad +\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+\nu_i)}{\sigma}\right)\right) \\
&\qquad\qquad \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{V_i|Y_i}(\nu_i)d\nu_i \\
&\qquad -\int_{\mathbb{R}}\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\right)f_{V_i|Y_i}(\nu_i)d\nu_i \\
&\qquad \left.\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{V_i|Y_i}(\nu_i)d\nu_i\right) \\
&= \sum_{i=1}^{D}\left(E\left[-\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}}-\tau)x_{ij}\middle|Y_i\right]\right. \\
&\qquad +E\left[\left(-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\right)\right. \\
&\qquad\qquad\qquad \left.\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}}-\tau)x_{ij}\right)\middle|Y_i\right] \\
&\qquad -E\left[-\frac{n_i}{\sigma}+\frac{1}{\sigma}\sum_{j=1}^{n_i}\rho_\tau\left(\frac{Y_{ij}-(x_{ij}^T\beta_\tau+V_i)}{\sigma}\right)\middle|Y_i\right] \\
&\qquad \left.E\left[-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+V_i\}}-\tau)x_{ij}\middle|Y_i\right]\right).
\end{aligned}
$$

Rewriting this result in vector notation I get

$$
\begin{aligned}
&-\frac{1}{\sigma^2}E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right] \\
&-\frac{1}{\sigma^3}\left(E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}\right.\right. \\
&\qquad\qquad \left.\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right] \\
&\qquad -E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right] \\
&\qquad\quad \left.E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T(Y-(\mathbf{X}\beta_\tau+\mathbf{Z}V))\,|Y\right]\right).
\end{aligned}
$$

Using the model (4.25) the derivative develops to

$$
\begin{aligned}
&-\frac{1}{\sigma^2}\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right]\\
&-\frac{1}{\sigma^3}\left(\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\right.\\
&\qquad\left.-\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right]\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\right)\\
&\quad=-\frac{1}{\sigma^2}\left(\mathrm{E}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right]-\tau\mathbf{1}_n^T\right)\mathbf{X}\\
&\quad-\frac{1}{\sigma^3}\left(\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T|Y\right]\mathbf{X}\right.\\
&\qquad\left.-\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T|Y\right]\mathbf{X}\right)\\
&\quad=-\frac{1}{\sigma^2}\left(\mathrm{E}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right]-\tau\mathbf{1}_n^T\right)\mathbf{X}\\
&\quad-\frac{1}{\sigma^3}\mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon,\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X},
\end{aligned}
$$

where $\mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon,\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)$ is the n-dimensional conditional covariance between the variable $\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon$ and the vector $\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}$. □

4.1.3.5.1 The Derivatives of $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$

Generally, the derivative of $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ with respect to θ_ι is given by

$$
\begin{aligned}
&\frac{\partial}{\partial\theta_\iota}\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)\\
&\quad=\frac{\partial}{\partial\theta_\iota}\left(\sum_{i=1}^{D}\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\right.\\
&\qquad\qquad\left.f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i\right)\\
&\quad\overset{(\star)}{=}\sum_{i=1}^{D}\left(\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\frac{\partial}{\partial\theta}\left(\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\right.\right.\\
&\qquad\qquad\left.f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)\right)d\nu_i\\
&\qquad-\frac{1}{f^2_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)f_{Y_i|V_i}(Y_i|\nu_i)\\
&\qquad\qquad\left.f_{V_i}(\nu_i)d\nu_i\frac{\partial}{\partial\theta_\iota}f_{Y_i}(Y_i)\right),
\end{aligned}
\tag{4.28}
$$

where $(\star)$ holds by applying the *Lebesgue Theorem*. The nominator in the first quotient can be simplified to

$$\int_{\mathbb{R}} \frac{\partial}{\partial\theta_\iota}\left(\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)\right) d\nu_i$$

$$= \int_{\mathbb{R}}\left(\frac{\partial}{\partial\theta_\iota}\left(\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{Y_i|V_i}(Y_i|\nu_i)\right) f_{V_i}(\nu_i) + \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{Y_i|V_i}(Y_i|\nu_i) \frac{\partial}{\partial\theta_\iota}f_{V_i}(\nu_i)\right) d\nu_i.$$

In this expression the first summand is zero, whenever $\theta_\iota \in \Theta_V$ and the second one is zero, whenever $\theta_\iota \in \Theta_Y$. The derivatives of $f_{Y_i}(Y_i)$ in the second quotient are given in (4.22), (4.23), and (4.24).
By application of the *Schwarz' Theorem* – see (4.20) – I have

$$\frac{\partial}{\partial\sigma_V}\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y) = \frac{\partial}{\partial\beta_\tau}\frac{\partial}{\partial\sigma_V}\ell(\theta|Y), \tag{4.29}$$

which is given in case 3 – see Lemma 4.11 – and

$$\frac{\partial}{\partial\sigma}\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y) = \frac{\partial}{\partial\beta_\tau}\frac{\partial}{\partial\sigma}\ell(\theta|Y), \tag{4.30}$$

which is given in case 5 – see Lemma 4.13. Thus the only derivative of $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ I have to calculate is with respect to β_τ itself, which is the last case.

Case 6: The Second Derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\beta_\tau, \beta_\tau)$

The sixth case is the second derivative of $\ell(\theta|Y)$ twice with respect to β_τ. I start with the first derivative $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ given in Lemma 4.7, which is a p-dimensional vector and differentiate it with respect to the p-dimensional parameter vector β_τ leading to a $p \times p$-dimensional matrix, which is given in the following lemma.

Lemma 4.14. *The second derivative of $\ell(\theta|Y)$ with respect to $(\theta_{\iota_1}, \theta_{\iota_2}) = (\beta_\tau, \beta_\tau)$ is given by*

$$\frac{\partial^2}{\partial\beta_\tau^2}\ell(\theta|Y) = -\frac{1}{\sigma^2}\mathbf{X}^T \mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}, \mathbf{1}_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}.$$

Proof. For the parameter β_τ the derivative of $\frac{\partial}{\partial\beta_\tau}\ell(\theta|Y)$ simplifies with (4.28) and previous argumentation to

$$\begin{aligned}
&\frac{\partial^2}{\partial\beta_\tau^2}\ell(\theta|Y)\\
&= \sum_{i=1}^{D}\left(\frac{1}{f_{Y_i}(Y_i)}\int_{\mathbb{R}}\frac{\partial}{\partial\beta_\tau}\left(\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\right.\right.\\
&\qquad\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i)\right) f_{V_i}(\nu_i)d\nu_i\\
&\qquad -\frac{1}{f^2_{Y_i}(Y_i)}\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\\
&\qquad\qquad\left. f_{Y_i|V_i}(Y_i|\nu_i)f_{V_i}(\nu_i)d\nu_i\frac{\partial}{\partial\beta_\tau}f_{Y_i}(Y_i)\right).
\end{aligned}$$

The derivative in the first integral can be calculated using (4.19) by

$$\begin{aligned}
&\frac{\partial}{\partial\beta_\tau}\left(\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{Y_i|V_i}(Y_i|\nu_i)\right)\\
&= \frac{\partial}{\partial\beta_\tau}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad + \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\frac{\partial}{\partial\beta_\tau}f_{Y_i|V_i}(Y_i|\nu_i)\\
&= 0\cdot f_{Y_i|V_i}(Y_i|\nu_i)\\
&\quad + \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\\
&\qquad \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)^T f_{Y_i|V_i}(Y_i|\nu_i)\\
&= \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)\\
&\quad \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbb{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)^T f_{Y_i|V_i}(Y_i|\nu_i).
\end{aligned}$$

Inserting this in the first equation of the proof leads to

$$
\begin{aligned}
&\frac{\partial^2}{\partial \beta_\tau^2}\ell(\theta|Y) \\
&= \sum_{i=1}^{D}\Bigg(\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) \\
&\qquad\qquad \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)^T f_{V_i|Y_i}(\nu_i)d\nu_i \\
&\quad -\left(\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{V_i|Y_i}(\nu_i)d\nu_i\right) \\
&\quad \left(\int_{\mathbb{R}}\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) f_{V_i|Y_i}(\nu_i)d\nu_i\right)^T\Bigg) \\
&= \sum_{i=1}^{D}\Bigg(E\Bigg[\left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right) \\
&\qquad\qquad \left(-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\right)^T \Bigg| Y_i\Bigg] \\
&\quad -E\left[-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\Bigg| Y_i\right] \\
&\quad E^T\left[-\frac{1}{\sigma}\sum_{j=1}^{n_i}(\mathbf{1}_{\{Y_{ij}\leqslant x_{ij}^T\beta_\tau+\nu_i\}}-\tau)x_{ij}\Bigg| Y_i\right]\Bigg).
\end{aligned}
$$

I can rewrite the second derivative in vector notation as follows

$$
\begin{aligned}
&-\frac{1}{\sigma^2}\Big(E\Big[\left(\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)\right)\left(\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)\right)^T|Y\Big] \\
&\quad -E\left[\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)|Y\right] \\
&\quad E^T\left[\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)|Y\right]\Big) \\
&= -\frac{1}{\sigma^2}\Big(E\left[\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right] \\
&\quad -E\left[\mathbf{X}^T\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)|Y\right] \\
&\quad E\left[\left(\mathbf{1}_{\{Y\leqslant \mathbf{X}\beta_\tau+\mathbf{Z}V\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right]\Big).
\end{aligned}
$$

By applying model (4.25) it holds

$$
\begin{aligned}
&-\frac{1}{\sigma^2}\Big(E\left[\mathbf{X}^T\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right] \\
&\quad -E\left[\mathbf{X}^T\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)|Y\right]E\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\mathbf{X}|Y\right]\Big)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{\sigma^2}\left(E\left[\mathbf{X}^T \mathbf{1}_{\{\varepsilon \leqslant 0\}} \mathbf{1}_{\{\varepsilon \leqslant 0\}}^T \mathbf{X}|Y\right]\right. \\
&\qquad \left. - E\left[\mathbf{X}^T \mathbf{1}_{\{\varepsilon \leqslant 0\}}|Y\right] E\left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T \mathbf{X}|Y\right]\right) \\
&= -\frac{1}{\sigma^2}\mathbf{X}^T\left(E\left[\mathbf{1}_{\{\varepsilon \leqslant 0\}} \mathbf{1}_{\{\varepsilon \leqslant 0\}}^T|Y\right]\right. \\
&\qquad \left. - E\left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}|Y\right] E\left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T|Y\right]\right)\mathbf{X} \\
&= -\frac{1}{\sigma^2}\mathbf{X}^T \mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}, \mathbf{1}_{\{\varepsilon \leqslant 0\}}|Y\right)\mathbf{X},
\end{aligned}
$$

where $\mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}, \mathbf{1}_{\{\varepsilon \leqslant 0\}}|Y\right)$ is the conditional $n \times n$-dimensional covariance matrix of the p-dimensional random vector variable $\mathbf{1}_{\{\varepsilon \leqslant 0\}}$.

□

4.2 THE PROOF OF THE ASSUMPTIONS IN THE WEISS' THEOREM

After I have calculated all second derivatives of the log-likelihood density, I am prepared for the proof of the two assumptions in the *Weiss' Theorem* – see Theorem 4.1. Since this procedure is quite complex, I will first show Assumption 1 in Section 4.2.1 followed by the proof of Assumption 2 in Section 4.2.2.

4.2.1 *Proof of Assumption 1*

For Assumption 1 as stated in (4.1) I need the limit in probability, when $n \to \infty$ of

$$-\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}}\ell(\theta|Y)$$

at the point θ^0, which represents the true but unknown parameter vector. The sequences $K_\iota(n)$ must be derived for each unknown parameter $\theta_\iota \in \{\sigma_V, \sigma, \beta_\tau\}$.

All second derivatives have the form of expectations, variances, and covariances conditional on Y. These expressions are random variables and can be seen as sums. By the *Law of Large Numbers* which can be applied with assumption (B5) and an appropriate choice for $K_\iota(n)$ they converge to their expected value. Thus for an expression $\xi(V, \varepsilon)$

$$\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}E\left[\xi(V,\varepsilon)|Y\right] \to E[E\left[\xi(V,\varepsilon)|Y\right]] \tag{4.31}$$

$$\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}Var\left(\xi(V,\varepsilon)|Y\right) \to E[Var\left(\xi(V,\varepsilon)|Y\right)] \tag{4.32}$$

and

$$\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)}\mathrm{Cov}\left(\xi_1(V,\varepsilon),\xi_2(V,\varepsilon)|Y\right) \to E[\mathrm{Cov}\left(\xi_1(V,\varepsilon),\xi_2(V,\varepsilon)|Y\right)]. \tag{4.33}$$

The expected value of a conditional expectation of an expression $\xi(V, \varepsilon)$ is its unconditional expectation

$$E[E[\xi(V, \varepsilon)|Y]] = E[\xi(V, \varepsilon)],$$

which can be seen in Satz 8.14 (iv) in Klenke [2013]. From Steyer [2003] p. 189 I also know that the expectations of conditional variances and covariances of expressions $\xi(V, \varepsilon)$ are the unconditional variances and covariances of $\xi(V, \varepsilon)$, where the variance and covariance of the conditional expectations is subtracted, respectively,

$$E[Var(\xi(V, \varepsilon)|Y)] = Var(\xi(V, \varepsilon)) - Var(E[\xi(V, \varepsilon)|Y])$$

and

$$\begin{aligned} E[Cov(\xi_1(V, \varepsilon), \xi_2(V, \varepsilon)|Y)] &= Cov(\xi_1(V, \varepsilon), \xi_2(V, \varepsilon)) \\ &\quad - Cov(E[\xi_1(V, \varepsilon)|Y], E[\xi_2(V, \varepsilon)|Y]). \end{aligned}$$

The appropriate choice of the convergence rates $K_\iota(n)$ will be discussed in the following Section 4.2.1.1. With the application of the *Law of Large Numbers* I get the limit of all second derivatives given as expected values, variances, and covariances, which will be calculated in Section 4.2.1.2.

4.2.1.1 *The Convergence Rates* $K(n)$

The convergence rates will be defined for each parameter θ_ι separately. They determine the speed of convergence of the corresponding parameter estimation. Their choice will detect the part of the sample size, D or n, which improves the estimation of the particular parameter. Thus the number of areas D will determine the accuracy of estimation of σ_V

$$K_1(n) = K_{\sigma_V} = \sqrt{D}.$$

As an additional assumption, D must grow of same order as n such that the *Law of Large Numbers* is applicable:

$$K_{\sigma_V} = \sqrt{D} = \mathcal{O}(n) \tag{4.34}$$

as $n \to \infty$ which is already assumed in (B1). This means that the number of samples within the areas does not increase faster than the number of areas D. For example, an increase of the sample size by keeping the number of areas stable is not covered by this asymptotics. Nevertheless the overall sample size n determines the convergence rates of σ and β_τ

$$K_\iota(n) = \sqrt{n} \quad \forall \iota = 2, 3.$$

By increase of D the overall sample size n as defines in (3.2) already rises, whenever the within area sample sizes n_i remain the same and are not zero in the new areas. By terms of accuracy of the estimation

of σ and β_τ, n increases even faster, whenever the within area sample sizes n_i increase, while respecting the assumption stated in (4.34) before.
Pinheiro [1994] showed the same convergence rates in the mean estimation in the linear mixed model.

4.2.1.2 *The Limit and Inverse Asymptotic Covariance Matrix* $B(\theta^0)$

The limit in Assumption 1 (4.1) will determine the entries of the inverse of the asymptotic covariance matrix. It is compounded by all limits of the second derivatives I have calculated in cases 1 to 6.

Lemma 4.15. *The inverse of the asymptotic covariance matrix is given by*

$$B(\theta^0) := \begin{pmatrix} B_{1,1}(\theta^0) & B_{1,2}(\theta^0) & B_{1,3}(\theta^0) \\ B_{2,1}(\theta^0) & B_{2,2}(\theta^0) & B_{2,3}(\theta^0) \\ B_{3,1}(\theta^0) & B_{3,2}(\theta^0) & B_{3,3}(\theta^0) \end{pmatrix},$$

where

$$\begin{aligned}
B_{1,1}(\theta^0) &= B_{\sigma_V,\sigma_V}(\theta^0) \\
&:= -\frac{1}{{\sigma_V^0}^2} + \frac{1}{{\sigma_V^0}^6} E\left[E^2\left[V_1^2|Y\right]\right]\Big|_{\theta^0} \\
B_{1,2}(\theta^0) &= B_{2,1}(\theta^0) = B_{\sigma_V,\sigma}(\theta^0) = B_{\sigma,\sigma_V}(\theta^0) := 0 \\
B_{1,3}^T(\theta^0) &= B_{3,1}(\theta^0) = B_{\sigma_V,\beta_\tau}^T(\theta^0) = B_{\beta_\tau,\sigma_V}(\theta^0) := 0_p \\
B_{2,2}(\theta^0) &= B_{\sigma,\sigma}(\theta^0) \\
&:= -\frac{3}{{\sigma^0}^2} - \frac{2\tau}{{\sigma^0}^2(1-\tau)^2} + \frac{1}{{\sigma^0}^2} E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]\Big|_{\theta^0} \\
B_{2,3}^T(\theta^0) &= B_{3,2}(\theta^0) = B_{\sigma,\beta_\tau}^T(\theta^0) = B_{\beta_\tau,\sigma}(\theta^0) \\
&:= \frac{1}{{\sigma^0}^3}\left(-\tau\sigma - E\left[E\left[(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau)\varepsilon_{1,1}|Y\right] E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\Big|_{\theta^0}\right) c_0 \\
B_{3,3}(\theta^0) &= B_{\beta_\tau,\beta_\tau}(\theta^0) \\
&:= -\frac{1}{{\sigma^0}^2}\left(\tau - E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\Big|_{\theta^0}\right) C_1
\end{aligned}$$

where c_0 *is a p-dimensional vector and* C_1 *is a* $p \times p$ *dimensional matrix (cf. Assumptions (B3) and (B4)).*

Proof. In this proof I show the convergence in Assumption 1 with the application of the *Law of Large Numbers* which can be applied with assumption (B5) in each of the six cases. There I will use the preliminaries stated in the beginning of this Section 4.2.1.

Case 1: By application of the *Law of Large Numbers* for $D \to \infty$ I get for $K_{\sigma_V}(n) = \sqrt{D} = \mathcal{O}(n)$

$$\frac{1}{D}E\left[V^T V|Y\right] = \frac{1}{D}\sum_{i=1}^{D} E\left[V_i^2|Y\right] \to E\left[E\left[V_1^2|Y\right]\right] = E\left[V_1^2\right] \overset{(\star)}{=} \sigma_V^2$$

where V_1 stands for a representative of the random effect V_i and follows its distribution and

$$\begin{aligned}
\frac{1}{D}Var\left(V^T V|Y\right) &= \frac{1}{D}\sum_{i=1}^{D} Var\left(V_i^2|Y\right) \\
&\to E\left[Var\left(V_1^2|Y\right)\right] = Var\left(V_1^2\right) - Var\left(E\left[V_1^2|Y\right]\right) \\
&= E\left[V_1^4\right] - E^2\left[V_1^2\right] - E\left[E^2\left[V_1^2|Y\right]\right] + E^2\left[E\left[V_1^2|Y\right]\right] \\
&= E\left[V_1^4\right] - E^2\left[V_1^2\right] - E\left[E^2\left[V_1^2|Y\right]\right] + E^2\left[V_1^2\right] \\
&\overset{(\star)}{=} 3\sigma_V^4 - E\left[E^2\left[V_1^2|Y\right]\right],
\end{aligned}$$

where $(\star)$ follows from $V_1 \sim N(0, \sigma_V^2)$. By Lemma 4.9 this leads to

$$\begin{aligned}
B_{\sigma_V,\sigma_V}(\theta) &= -\underset{D\to\infty}{\text{P-lim}}\frac{1}{D}\left(\frac{D}{\sigma_V^2} - \frac{3}{\sigma_V^4}E\left[V^T V|Y\right] + \frac{1}{\sigma_V^6}Var\left(V^T V|Y\right)\right) \\
&= -\frac{1}{\sigma_V^2} + \frac{3}{\sigma_V^4}\sigma_V^2 - \frac{1}{\sigma_V^6}\left(3\sigma_V^4 - E\left[E^2\left[V_1^2|Y\right]\right]\right) \\
&= \frac{-1+3-3}{\sigma_V^2} + \frac{1}{\sigma_V^6}E\left[E^2\left[V_1^2|Y\right]\right] \\
&= -\frac{1}{\sigma_V^2} + \frac{1}{\sigma_V^6}E\left[E^2\left[V_1^2|Y\right]\right].
\end{aligned}$$

Case 2: By application of the *Law of Large Numbers* for $D \to \infty$ and $n \to \infty$ I get for $K_{\sigma_V}(n) = \sqrt{D}$ and $K_\sigma(n) = \sqrt{n}$

$$\begin{aligned}
&\frac{1}{\sqrt{D}\sqrt{n}}Cov\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T\varepsilon, V^T V|Y\right) \\
&\quad= \frac{1}{\sqrt{D}\sqrt{n}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} Cov\left(\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}, V_i^2|Y\right) \\
&\quad\to E\left[Cov\left(\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}, V_1^2|Y\right)\right] \\
&\quad\quad= Cov\left(\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}, V_1^2\right) \\
&\quad\quad\quad - Cov\left(E\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right], E\left[V_1^2|Y\right]\right) \\
&\quad\quad\overset{(\star)}{=} 0,
\end{aligned}$$

where $(\star)$ holds because of the independence of $E[\varepsilon_{1,1}|Y]$ and $E[V_1|Y]$ assumed in (B6). By Lemma 4.10 this leads to

$$\begin{aligned}
B_{\sigma_V,\sigma}(\theta) &= B_{\sigma,\sigma_V}(\theta) \\
&= -\operatorname*{P-lim}_{n,D\to\infty} \frac{1}{\sqrt{D}\sqrt{n}} \left(\frac{1}{\sigma^2\sigma_V^3} \mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau \mathbf{1}_n\right)^T \varepsilon, V^T V|Y\right)\right) \\
&= 0.
\end{aligned}$$

Case 3: By application of the *Law of Large Numbers* for $D \to \infty$ and $n \to \infty$ and (B3) I get for $K_{\sigma_V}(n) = \sqrt{D}$ and $K_{\beta_\tau}(n) = \sqrt{n}$

$$\begin{aligned}
&\frac{1}{\sqrt{D}\sqrt{n}} \mathrm{Cov}\left(V^T V, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right) \mathbf{X} \\
&= \frac{1}{\sqrt{D}\sqrt{n}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \mathrm{Cov}\left(V_i^2, \mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau|Y\right) x_{ij} \\
&\to \left(\mathrm{Cov}\left(V_1^2, \mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right) - \mathrm{Cov}\left(E\left[V_1^2|Y\right], E\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right)\right) c_0 \\
&\overset{(\star)}{=} 0_p,
\end{aligned}$$

where c_0 is a p-dimensional vector which is of form $E[X_{1,1}]$ in which $X_{1,1}$ is a representative of the covariate x (cf. assumption (B3)). 0_p is a p-dimensional vector of zeros and $(\star)$ holds because of the independence of $E[\varepsilon_{1,1}|Y]$ and $E[V_1|Y]$ stated in (B6). By Lemma 4.11 this leads to

$$\begin{aligned}
B^T_{\sigma_V,\beta_\tau}(\theta) &= B_{\beta_\tau,\sigma_V}(\theta) \\
&= -\operatorname*{P-lim}_{n\to\infty} \frac{1}{\sqrt{D}\sqrt{n}} \left(-\frac{1}{\sigma\sigma_V^3} \mathrm{Cov}\left(V^T V, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right) \mathbf{X}\right) \\
&= 0_p.
\end{aligned}$$

Case 4: By application of the *Law of Large Numbers* for $n \to \infty$ I get for $K_\sigma(n) = \sqrt{n}$

$$\begin{aligned}
\frac{1}{n} E\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] &= \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right] \\
&\to E\left[E\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right] = E\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right] \overset{(\star)}{=} -\frac{\tau\sigma}{1-\tau}, \\
\frac{1}{n}\mathbf{1}_n^T E\left[\varepsilon|Y\right] &= \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E\left[\varepsilon_{ij}|Y\right] \\
&\to E\left[E\left[\varepsilon_{1,1}|Y\right]\right] = E\left[\varepsilon_{1,1}\right] \overset{(\star)}{=} \frac{\sigma(1-2\tau)}{\tau(1-\tau)},
\end{aligned}$$

and

$$\frac{1}{n} \mathrm{Var}\left(\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right) = \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \mathrm{Var}\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right)$$

$$
\begin{aligned}
&\to E\left[Var\left(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right)\right]\\
&\quad = Var\left(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right) - Var\left(E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right)\\
&\quad = E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}^2\right] - E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right]\\
&\qquad - E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right] - E^2\left[E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]\\
&\quad = E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}^2\right] - E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right]\\
&\qquad - E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right] - E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right]\\
&\quad \overset{(\star)}{=} \frac{2\tau\sigma^2}{(1-\tau)^2} - E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right],
\end{aligned}
$$

where $(\star)$ follows from $\varepsilon_{1,1} \sim ALD(0, \sigma, \tau)$ – see Corollary 2.2 in Section 2.4.1 for details. By Lemma 4.12 this leads to

$$
\begin{aligned}
B_{\sigma,\sigma}(\theta) &= -\underset{n\to\infty}{\text{P-lim}}\,\frac{1}{n}\left(\frac{n}{\sigma^2} - \frac{2}{\sigma^3}\left(E\left[\mathbb{1}_{\{\varepsilon\leqslant 0\}}^T\varepsilon|Y\right] - \tau\mathbb{1}_n^T E\left[\varepsilon|Y\right]\right)\right.\\
&\qquad\left. + \frac{1}{\sigma^4}Var\left(\mathbb{1}_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right)\right)\\
&= -\frac{1}{\sigma^2} + \frac{2}{\sigma^3}\left(-\frac{\tau\sigma}{1-\tau} - \tau\frac{\sigma(1-2\tau)}{\tau(1-\tau)}\right)\\
&\quad - \frac{1}{\sigma^4}\left(\frac{2\tau\sigma^2}{(1-\tau)^2} - E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]\right)\\
&= -\frac{1}{\sigma^2} - \frac{2}{\sigma^2}\frac{1-2\tau+\tau}{1-\tau}\\
&\quad - \frac{2\tau}{\sigma^2(1-\tau)^2} + \frac{1}{\sigma^4}E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]\\
&= -\frac{3}{\sigma^2} - \frac{2\tau}{\sigma^2(1-\tau)^2} + \frac{1}{\sigma^4}E\left[E^2\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right].
\end{aligned}
$$

Case 5: By application of the *Law of Large Numbers* for $n \to \infty$ and (B3) I get for $K_\sigma(n) = K_{\beta_\tau}(n) = \sqrt{n}$

$$
\begin{aligned}
&\frac{1}{n}E\left[\mathbb{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\mathbf{X} = \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]x_{ij}\\
&\quad\to E\left[E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]E\left[X_{1,1}\right] = E\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right]c_0\\
&\quad = P\left(\varepsilon_{1,1}\leqslant 0\right)c_0 \overset{(\star)}{=} \tau c_0
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n}Cov\left(\left(\mathbb{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbb{1}_n\right)^T\varepsilon, \mathbb{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right)\mathbf{X}\\
&\quad = \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}Cov\left(\left(\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}, \mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right)x_{ij}\\
&\quad\to E\left[Cov\left(\left(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}, \mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right)\right]c_0
\end{aligned}
$$

$$
\begin{aligned}
&= \Big(\mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}, \mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right) \\
&\quad -\mathrm{Cov}\left(\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right], \mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right)\Big) c_0 \\
&= \Big(\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right] \\
&\quad -\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right] \\
&\quad -\mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right] \\
&\quad +\mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right]\right]\mathrm{E}\left[\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\Big) c_0 \\
&= \Big((1-\tau)\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}\right] \\
&\quad -\mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\Big) c_0 \\
&\overset{(\star)}{=} \left(-\tau\sigma - \mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}|Y\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\right) c_0,
\end{aligned}
$$

where c_0 is defined as above in case 3 and $(\star)$ follows from $\varepsilon_{1,1} \sim \mathrm{ALD}(0,\sigma,\tau)$ – see Corollary 2.2 in Section 2.4.1 for details. By Lemma 4.13 this leads to

$$
\begin{aligned}
B^T_{\sigma,\beta_\tau}(\theta) &= B_{\beta_\tau,\sigma}(\theta) \\
&= -\operatorname*{P\text{-}lim}_{n\to\infty}\frac{1}{n}\Big(-\frac{1}{\sigma^2}\left(\mathrm{E}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right] - \tau\mathbf{1}_n^T\right)\mathbf{X} \\
&\qquad -\frac{1}{\sigma^3}\mathrm{Cov}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T\varepsilon, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}\Big) \\
&= \frac{1}{\sigma^2}\left(\tau c_0 - \tau c_0\right) \\
&\quad + \frac{1}{\sigma^3}\left(-\tau\sigma - \mathrm{E}\left[\mathrm{E}\left[(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau)\varepsilon_{1,1}|Y\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\right) c_0 \\
&= \frac{1}{\sigma^3}\left(-\tau\sigma - \mathrm{E}\left[\mathrm{E}\left[(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau)\varepsilon_{1,1}|Y\right]\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\right) c_0.
\end{aligned}
$$

Case 6: By application of the *Law of Large Numbers* for $n \to \infty$ and (B4) I get for $K_{\beta_\tau}(n) = \sqrt{n}$

$$
\begin{aligned}
&\frac{1}{n}\mathbf{X}^T\mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}, \mathbf{1}_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X} \\
&\quad = \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}\mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}, \mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right)x_{ij}^T \\
&\quad \to \mathrm{E}\left[\mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}, \mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right)\right] C_1 \\
&\qquad = \mathrm{E}\left[\mathrm{Var}\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right)\right] C_1 \\
&\qquad = \left(\mathrm{Var}\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right) - \mathrm{Var}\left(\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right)\right) C_1 \\
&\qquad = \Big(\mathrm{E}\left[\mathbf{1}^2_{\{\varepsilon_{1,1}\leqslant 0\}}\right] - \mathrm{E}^2\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right] - \mathrm{E}\left[\mathrm{E}^2\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right] \\
&\qquad\qquad +\mathrm{E}^2\left[\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\Big) C_1
\end{aligned}
$$

$$= \left(\mathrm{E}\left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}}\right] - \mathrm{E}\left[\mathrm{E}^2\left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}}|Y\right]\right]\right) C_1$$
$$\overset{(\star)}{=} \left(\tau - \mathrm{E}\left[\mathrm{E}^2\left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}}|Y\right]\right]\right) C_1,$$

where C_1 is a $p \times p$-dimensional matrix which is of form $\mathrm{E}\left[X_{1,1}X_{1,1}^T\right]$ in which $X_{1,1}$ is a representative of the covariate x (cf. assumption (B4)) and $(\star)$ follows from $\varepsilon_{1,1} \sim \mathrm{ALD}(0, \sigma, \tau)$ – see Corollary 2.2 in Section 2.4.1 for details. By Lemma 4.14 this leads to

$$\begin{aligned} & B_{\beta_\tau,\beta_\tau}(\theta) \\ &= -\underset{n\to\infty}{\text{P-}\lim}\, \frac{1}{n}\left(\frac{1}{\sigma^2}\mathbf{X}^T \mathrm{Cov}\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}, \mathbf{1}_{\{\varepsilon \leqslant 0\}}|Y\right)\mathbf{X}^T\right) \\ &= -\frac{1}{\sigma^2}\left(\tau - \mathrm{E}\left[\mathrm{E}^2\left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}}|Y\right]\right]\right) C_1. \end{aligned}$$

□

4.2.1.3 *The Continuity of* $B(\theta)$ *in* θ^0

All entries of B_{ι_1,ι_2} as they are derived in Lemma 4.15 are combinations of continuous functions and integrals. The latter come from the expected values of the squared conditional expectations, which are not continuous only at

$$\{(\sigma_V, \sigma, \beta_\tau)|\varepsilon_{1,1} = 0\}.$$

Since this is a null set with respect to any continuous measure P, the expected values are P-almost surely (P-a.s.) continuous transformations. Altogether $B_{\iota_1,\iota_2}(\theta^0)$ is continuous for all $(\iota_1, \iota_2) \in \{1,2,3\} \times \{1,2,3\}$ and thus the whole matrix, as a linear function of P-almost surely continuous functions, is continuous.

4.2.1.4 *The Positive Definiteness of* $B(\theta^0)$

For the positive definiteness I will first state the definition (cf. Strang [2016]) and later prove this property for $B(\theta^0)$ as derived in Lemma 4.15.

Definition 4.16. *A symmetric matrix* $C \in \mathbb{R}^{k\times k}$ *is positive definite if and only if*

$$a^T C a > 0 \quad \forall a \in \mathbb{R}^k, a \neq 0_k. \tag{4.35}$$

The matrix $B(\theta^0)$ is symmetric and has similar properties to a covariance matrix. Therefore I can state the following lemma.

Lemma 4.17. $B(\theta^0)$ *as defined in Lemma 4.15 is positive definite.*

Proof. All entries of $B(\theta^0)$ have the form

$$B_{\iota_1,\iota_2}(\theta^0) = -E\left[\xi_{\theta_{\iota_1}}(V,\varepsilon)\xi_{\theta_{\iota_2}}(V,\varepsilon) + \xi_{\theta_{\iota_1},\theta_{\iota_2}}(V,\varepsilon)\right] + E\left[E\left[\xi_{\theta_{\iota_1}}(V,\varepsilon)|Y\right] E\left[\xi_{\theta_{\iota_2}}(V,\varepsilon)|Y\right]\right]$$

for all $\iota_1, \iota_2 = 1,2,3$, where $\xi_{\theta_{\iota_1}}$, $\xi_{\theta_{\iota_2}}$, and $\xi_{\theta_{\iota_1},\theta_{\iota_2}}$ are P-a.s. continuous functions. Note that $\xi_{\theta_{\iota_1},\theta_{\iota_2}}$ is zero, whenever θ_{ι_1} and θ_{ι_2} are not in the same subset Θ_Y or Θ_V. Thus

$$B(\theta^0) = -E\left[\xi_\theta(V,\varepsilon)\xi_\theta^T(V,\varepsilon) + \xi_{\theta,\theta}(V,\varepsilon)\right]\Big|_{\theta^0} + E\left[E\left[\xi_\theta(V,\varepsilon)|Y\right] E^T\left[\xi_\theta(V,\varepsilon)|Y\right]\right]\Big|_{\theta^0},$$

where $\xi_\theta := (\xi_{\sigma_V}, \xi_\sigma, \xi_{\beta_\tau}^T)^T$ and $\xi_{\theta,\theta}$ is the diagonal matrix with the entries $\xi_{\theta_{\iota_1},\theta_{\iota_2}}$. The expectation of a matrix is here defined componentwise. By *maximum likelihood* theory the entry of the first expectation comes from the second derivative of the density of Y and is therefore in expectation equal to zero, which leads for an arbitrary and fixed $a \in \mathbb{R}^k$ to

$$a^T B(\theta^0) a = a^T E\left[E\left[\xi_\theta(V,\varepsilon)|Y\right] E^T\left[\xi_\theta(V,\varepsilon)|Y\right]\right]\Big|_{\theta^0} a = E\left[a^T E\left[\xi_\theta(V,\varepsilon)|Y\right] E^T\left[\xi_\theta(V,\varepsilon)|Y\right] a\right]\Big|_{\theta^0}.$$

Note that

$$a^T E\left[\xi_\theta(V,\varepsilon)|Y\right] = E^T\left[\xi_\theta(V,\varepsilon)|Y\right] a,$$

which is a one-dimensional random variable and thus

$$a^T E_{V_1|Y_1}[\xi_\theta] E^T_{V_1|Y_1}[\xi_\theta] a \geqslant 0$$

as a squared real valued random variable. Then the expected value of this positive random variable is positive.

$$a^T B(\theta^0) a = E\left[a^T E\left[\xi_\theta(V,\varepsilon)|Y\right] E^T\left[\xi_\theta(V,\varepsilon)|Y\right] a\right]\Big|_{\theta^0} \geqslant 0,$$

which proves the positive semi-definiteness. Furthermore the matrix $E\left[\xi_\theta(V,\varepsilon)|Y\right]$ has linearly independent rows (see Lemma 4.15). Therefore it holds for all $a \neq 0_k$

$$a^T E_{V_1|Y_1}[\xi_\theta] E^T_{V_1|Y_1}[\xi_\theta] a \neq 0.$$

Altogether it follows

$$a^T B(\theta^0) a = E\left[a^T E\left[\xi_\theta(V,\varepsilon)|Y\right] E^T\left[\xi_\theta(V,\varepsilon)|Y\right] a\right]\Big|_{\theta^0} | > 0,$$

which proves the positive definiteness.

□

4.2.2 *Proof of Assumption 2*

In Section 4.2.1 I have shown that the second derivatives converge in all six cases to continuous functions. Their limits build a positive definite matrix $B(\theta)$, whose inverse turns out to be the asymptotic covariance matrix of $diag(K(n))(\hat{\theta}(n) - \theta^0)$. Assumption 2 also deals with this convergence but in terms of convergence speed. Thus it turns out that for the environment $N_n(\theta^0)$ around the true parameter θ^0 – as defined in (4.3) – the convergence will be faster than a given rate $M(n)$. The proof depends on the choice of $M(n)$ and is for that reason a matter of construction. In order to keep this part clearly arranged it will be executed for each of the six different second derivatives separately keeping their case numbers as they have been introduced before.

4.2.2.1 *The Sequence* $M(n)$

As mentioned before, the choice of the sequence $M(n)$ is a matter of construction. There are two main features of the sequence stated as assumptions in Theorem 4.1

$$M_\iota(n) \to \infty \quad \text{and} \quad \frac{M_\iota(n)}{K_\iota(n)} \to 0 \quad \text{as } n \to \infty. \tag{4.36}$$

Now let me define, similar to the approach in Pinheiro [1994],

$$\kappa(n) := \max_{\iota_1,\iota_2} \left| -\frac{1}{K_{\iota_1}(n)K_{\iota_2}(n)} E_{Y|\theta^0}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^0}\right] - B_{\iota_1,\iota_2}(\theta^0)\right|. \tag{4.37}$$

Then for all $\iota = 1, 2, 3$ define

$$M_\iota(n) := \min\left\{K_{\sigma_V}^{\frac{1}{4}}(n), K_\sigma^{\frac{1}{4}}(n), K_{\beta_\tau}^{\frac{1}{4}}(n), \kappa^{-\frac{1}{4}}(n)\right\}. \tag{4.38}$$

Thus the ι may be dropped. For reasons of convenience I will also drop the n in K_ι, M_ι, and κ. Both assumptions on M as restated in (4.36) are fulfilled by this definition: M is positive and it holds

$$M \geqslant K_{\sigma_V}^{\frac{1}{4}} = D^{\frac{1}{8}} \to \infty \quad \text{and}$$
$$\frac{M}{K_\iota} \leqslant \frac{K_\iota^{\frac{1}{4}}}{K_\iota} = K_\iota^{-\frac{3}{4}} \to 0, \quad \iota = 1,2,3$$

by assumption (4.34) as $n \to \infty$.

4.2.2.2 *Approach of Proofing the Convergence*

Already Miller [1977] and Pinheiro [1994] used the fact that in order to show Assumption 2 it is sufficient to show

$$M^2 \sup_{\theta^1 \in N_n(\theta^0)} \left| -\frac{1}{K_{\iota_1} K_{\iota_2}} \frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^1} - B_{\iota_1,\iota_2}(\theta^0) \right| \overset{P_{Y|\theta^2}}{\to} 0$$
$$\forall \theta^2 \in N_n(\theta^0), \iota_1, \iota_2 \in \{1,2,\dots,k\}.$$

Following a similar approach as in Pinheiro [1994] (cf. (3.1.4)) the difference may be extended by adding zeros as follows

$$\begin{aligned}
&\sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{K_{\iota_1} K_{\iota_2}} \frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^1} - B_{\iota_1,\iota_2}(\theta^0) \right) \\
&= \underbrace{\sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^1} - \frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^2} \right) \right)}_{=:\phi_1} \\
&\quad \underbrace{- \frac{1}{K_{\iota_1} K_{\iota_2}} \left(\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^2} - E_{\theta^2}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^2} \right] \right)}_{=:\phi_2} \\
&\quad \underbrace{- \frac{1}{K_{\iota_1} K_{\iota_2}} \left(E_{\theta^2}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^2} \right] - E_{\theta^2}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^0} \right] \right)}_{=:\phi_3} \\
&\quad \underbrace{- \frac{1}{K_{\iota_1} K_{\iota_2}} \left(E_{\theta^2}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^0} \right] - E_{\theta^0}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^0} \right] \right)}_{=:\phi_4} \\
&\quad \underbrace{- \frac{1}{K_{\iota_1} K_{\iota_2}} E_{\theta^0}\left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\Big|_{\theta^0} \right] - B_{\iota_1,\iota_2}(\theta^0)}_{=:\phi_5},
\end{aligned}$$

where $\theta^1, \theta^2 \in N_n(\theta^0)$. At the same time the preserved sums are split into smaller parts labelled by ϕ_1, ϕ_2, ..., ϕ_5. E_θ denotes the expectation under the measure $P_{Y|\theta}$ conditional on the parameter θ. In the following I am going to show the convergence in probability of the defined ϕ's to zero for each of the six cases. Most of them have expressions like

$$\frac{1}{K_{\iota_1} K_{\iota_2}} E_\theta \left[\frac{\partial^2}{\partial\theta_{\iota_1}\partial\theta_{\iota_2}} \ell(\theta|Y)\bigg|_{\theta^\star} \right], \tag{4.39}$$

where $\theta, \theta^\star \in N_n(\theta^0)$. The expectation in (4.39) is for each case derived in the beginning of the corresponding Section. As a remark I have to state that $\theta^0 = (\sigma_V^0, \sigma^0, \beta_\tau^{0\,T})^T$. Furthermore θ^1 and θ^2 are defined in a similar fashion. This may create confusion, especially, when I write σ_V^2 for example. In the rest of this Section all parameter always have an index. Thus σ_V^2 is the second parameter from θ^2. The squared of σ_V^2 is denoted by $\sigma_V^{2\,2}$.
In the following the convergence is often shown in two steps. In the first step θ^1 and θ^2 are fixed while the expectations (4.39) converge with $n \to \infty$. In the second step it is used that the set $N_n(\theta^0)$ is shrinking (at the same time).
The *Law of Large Numbers* is always applied employing assumption (B5). As discussed before this is a commonly used regularity condition in quantile regression.

4.2.2.3 *Case 1*

The expectation (4.39) can be expressed by

$$\begin{aligned} &\frac{1}{D} E_\theta \left[\frac{\partial^2}{\partial\sigma_V^2} \ell(\theta|Y)\bigg|_{\theta^\star} \right] \\ &\quad = \frac{1}{D} E_\theta \left[\frac{D}{\sigma_V^{\star\,2}} - \frac{3}{\sigma_V^{\star\,4}} E_{\theta^\star}\left[V^T V|Y\right] + \frac{1}{\sigma_V^{\star\,6}} Var_{\theta^\star}\left(V^T V|Y\right) \right]. \end{aligned}$$

The variance may be split as follows

$$\begin{aligned} &\frac{1}{D} E_\theta \left[Var_{\theta^\star}\left(V^T V|Y\right) \right] \\ &\quad = \frac{1}{D} \left(E_\theta \left[E_{\theta^\star}\left[V^T V V^T V|Y\right] \right] - E_\theta \left[E_{\theta^\star}^2 \left[V^T V|Y\right] \right] \right), \end{aligned}$$

which leads to

$$\begin{aligned} &\frac{1}{D} E_\theta \left[\frac{\partial^2}{\partial\sigma_V^2} \ell(\theta|Y)\bigg|_{\theta^\star} \right] \\ &\quad = \frac{1}{\sigma_V^{\star\,2}} - \frac{3}{D\sigma_V^{\star\,4}} E_\theta \left[E_{\theta^\star}\left[V^T V|Y\right] \right] \\ &\qquad + \frac{1}{D\sigma_V^{\star\,6}} \left(E_\theta \left[E_{\theta^\star}\left[V^T V V^T V|Y\right] \right] - E_\theta \left[E_{\theta^\star}^2 \left[V^T V|Y\right] \right] \right) \end{aligned}$$

$$= \frac{1}{{\sigma_V^\star}^2} - \frac{3}{D{\sigma_V^\star}^4} \sum_{i=1}^{D} E_\theta \left[E_{\theta^\star} \left[V_i^2 | Y \right] \right]$$
$$+ \frac{1}{D{\sigma_V^\star}^6} \left(\sum_{i=1}^{D} E_\theta \left[E_{\theta^\star} \left[V_i^4 | Y \right] \right] - \sum_{i=1}^{D} E_\theta \left[E_{\theta^\star}^2 \left[V_i^2 | Y \right] \right] \right). \quad (4.40)$$

Whenever $\theta = \theta^\star$ this expression simplifies with $V_i \overset{iid}{\sim} N(0, {\sigma_V^\star}^2)$ and $E_{\theta^\star}[E_{\theta^\star}[\cdot|Y]] = E_{\theta^\star}[\cdot]$ to

$$\frac{1}{D} E_{\theta^\star} \left[\frac{\partial^2}{\partial \sigma_V^2} \ell(\theta|Y) \Big|_{\theta^\star} \right]$$
$$= \frac{1}{{\sigma_V^\star}^2} - \frac{3}{D{\sigma_V^\star}^4} \sum_{i=1}^{D} E_{\theta^\star} \left[V_i^2 \right]$$
$$+ \frac{1}{D{\sigma_V^\star}^6} \sum_{i=1}^{D} E_{\theta^\star} \left[V_i^4 \right] - \frac{1}{D{\sigma_V^\star}^6} \sum_{i=1}^{D} E_{\theta^\star} \left[E_{\theta^\star}^2 \left[V_i^2 | Y \right] \right]$$
$$= \frac{1}{{\sigma_V^\star}^2} - \frac{3}{{\sigma_V^\star}^4} {\sigma_V^\star}^2 + \frac{3}{{\sigma_V^\star}^6} {\sigma_V^\star}^4 - \frac{1}{D{\sigma_V^\star}^6} \sum_{i=1}^{D} E_{\theta^\star} \left[E_{\theta^\star}^2 \left[V_i^2 | Y \right] \right]$$
$$= \frac{1}{{\sigma_V^\star}^2} - \frac{1}{D{\sigma_V^\star}^6} \sum_{i=1}^{D} E_{\theta^\star} \left[E_{\theta^\star}^2 \left[V_i^2 | Y \right] \right]. \quad (4.41)$$

4.2.2.3.1 Case 1: Convergence of ϕ_1

For ϕ_1 in case 1 I have by application of Lemma 4.9 the expression

$$M^2 \phi_1 = M^2 \sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{D} \left(\frac{D}{{\sigma_V^1}^2} - \frac{3}{{\sigma_V^1}^4} E_{\theta^1} \left[V^T V | Y \right] \right.\right.$$
$$+ \frac{1}{{\sigma_V^1}^6} Var_{\theta^1} \left(V^T V | Y \right) - \frac{D}{{\sigma_V^2}^2}$$
$$\left.\left. + \frac{3}{{\sigma_V^2}^4} E_{\theta^2} \left[V^T V | Y \right] - \frac{1}{{\sigma_V^2}^6} Var_{\theta^2} \left(V^T V | Y \right) \right) \right)$$
$$= M^2 \sup_{\theta^1 \in N_n(\theta^0)} \left(- \left(\frac{1}{{\sigma_V^1}^2} - \frac{1}{{\sigma_V^2}^2} \right) \right.$$
$$- \frac{1}{D} \sum_{i=1}^{D} \left(\frac{3E_{\theta^1} \left[V_i^2 | Y \right]}{{\sigma_V^1}^4} - \frac{3E_{\theta^2} \left[V_i^2 | Y \right]}{{\sigma_V^2}^4} \right)$$
$$+ \frac{1}{D} \sum_{i=1}^{D} \left(\frac{E_{\theta^1} \left[V_i^4 | Y \right]}{{\sigma_V^1}^6} - \frac{E_{\theta^2} \left[V_i^4 | Y \right]}{{\sigma_V^2}^6} \right)$$
$$\left. - \frac{1}{D} \sum_{i=1}^{D} \left(\frac{E_{\theta^1}^2 \left[V_i^2 | Y \right]}{{\sigma_V^1}^6} - \frac{E_{\theta^2}^2 \left[V_i^2 | Y \right]}{{\sigma_V^2}^6} \right) \right),$$

where each summand converges to zero as $n \to \infty$ as shown in the following: By applying the *binomial formula* I have for the first summand the inequality

$$\frac{1}{{\sigma_V^1}^2} - \frac{1}{{\sigma_V^2}^2} = \frac{{\sigma_V^2}^2 - {\sigma_V^1}^2}{{\sigma_V^1}^2 {\sigma_V^2}^2} = \frac{\left(\sigma_V^2 - \sigma_V^1\right)\left(\sigma_V^2 + \sigma_V^1\right)}{{\sigma_V^1}^2 {\sigma_V^2}^2}.$$

I know that $\sigma_V^1, \sigma_V^2 \in N_{1,n}(\theta^0)$, where $N_{1,n}$ is the projection on first dimension of the $(p+2)$-dimensional set N_n and in this set holds

$$\left|\sigma_V^1 - \sigma_V^0\right| \leqslant \frac{M}{K_1} \quad \text{and} \quad \left|\sigma_V^2 - \sigma_V^0\right| \leqslant \frac{M}{K_1}.$$

Especially the first statement holds for all $\sigma_V^1 \in N_{1,n}(\theta^0)$ and therefore also for the supremum

$$\sup_{\sigma_V^1 \in N_{1,n}(\theta^0)} \left|\sigma_V^1 - \sigma_V^0\right| \leqslant \frac{M}{K_1}.$$

It follows by the *triangle inequality* that

$$\sup_{\sigma_V^1 \in N_{1,n}(\theta^0)} \left|\sigma_V^2 - \sigma_V^1\right| \leqslant \left|\sigma_V^2 - \sigma_V^0\right| + \sup_{\sigma_V^1 \in N_{1,n}(\theta^0)} \left|\sigma_V^1 - \sigma_V^0\right| \leqslant 2\frac{M_1}{K_1}. \tag{4.42}$$

By the *Law of Large Numbers* $\frac{1}{D}\sum_{i=1}^{D} E_{\theta^1}\left[V_i^2|Y\right]$ and $\frac{1}{D}\sum_{i=1}^{D} E_{\theta^2}\left[V_i^2|Y\right]$ converge as $n \to \infty$ and thus $D \to \infty$ to $E_{\theta^2}\left[V_1^2\right] = {\sigma_V^2}^2$ and $E_{\theta^1}\left[V_1^2\right] = {\sigma_V^1}^2$, respectively. Hence the second summand converges to

$$\frac{1}{D}\left(\frac{3E_{\theta^1}\left[V^T V|Y\right]}{{\sigma_V^1}^4} - \frac{3E_{\theta^2}\left[V^T V|Y\right]}{{\sigma_V^2}^4}\right)$$
$$\to \frac{3{\sigma_V^1}^2}{{\sigma_V^1}^4} - \frac{3{\sigma_V^2}^2}{{\sigma_V^2}^4} = 3\left(\frac{1}{{\sigma_V^1}^2} - \frac{1}{{\sigma_V^2}^2}\right).$$

As mentioned before, note that for the convergence σ_V^1 and σ_V^2 are fixed. Since $\sigma_V^1, \sigma_V^2 \in N_n(\theta^0)$ it holds for the limit that it is by (4.42) bounded from above by $\frac{M}{K_1}$ times a constant. In the third summand I can apply the *Law of Large Numbers*, again. It holds that $\frac{1}{D}\sum_{i=1}^{D} E_{\theta^1}\left[V_i^4|Y\right]$ and $\frac{1}{D}\sum_{i=1}^{D} E_{\theta^2}\left[V_i^4|Y\right])$ converge as $n \to \infty$ to $E_{\theta^1}\left[V_i^4\right] = 3{\sigma_V^1}^4$ and $E_{\theta^2}\left[V_i^4\right] = 3{\sigma_V^2}^4$, respectively. Thus

$$\frac{1}{D}\left(\frac{E_{\theta^1}\left[V_i^4|Y\right]}{{\sigma_V^1}^6} - \frac{E_{\theta^2}\left[V_i^4|Y\right]}{{\sigma_V^2}^6}\right) \to \frac{3{\sigma_V^1}^4}{{\sigma_V^1}^6} - \frac{3{\sigma_V^2}^4}{{\sigma_V^2}^6} = 3\left(\frac{1}{{\sigma_V^1}^2} - \frac{1}{{\sigma_V^2}^2}\right),$$

which is by (4.42) bounded from above by $\frac{M}{K_1}$ times a constant. By the definition of $N_n(\theta^0)$ the random variables $E_{\theta^1}^2\left[V_i^2|Y\right]$ and $E_{\theta^2}^2\left[V_i^2|Y\right]$

converge as $n \to \infty$ to $E^2_{\theta^0}\left[V_i^2|Y\right]$. At the same time by the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variables converges to its expectation and thus for the random variables $E^2_{\theta^1}\left[V_i^2|Y\right]$ and $E^2_{\theta^2}\left[V_i^2|Y\right]$ I get

$$\frac{1}{D}\sum_{i=1}^{D} E^2_{\theta^1}\left[V_i^2|Y\right] \to E_{\theta^0}\left[E^2_{\theta^0}\left[V_1^2|Y\right]\right]$$

and

$$\frac{1}{D}\sum_{i=1}^{D} E^2_{\theta^2}\left[V_i^2|Y\right] \to E_{\theta^0}\left[E^2_{\theta^0}\left[V_1^2|Y\right]\right].$$

Hence I have for the fourth summand

$$\frac{1}{D}\sum_{i=1}^{D}\left(\frac{E^2_{\theta^1}\left[V_i^2|Y\right]}{{\sigma_V^1}^6} - \frac{E^2_{\theta^2}\left[V_i^2|Y\right]}{{\sigma_V^2}^6}\right) \to \left(\frac{1}{{\sigma_V^1}^6} - \frac{1}{{\sigma_V^2}^6}\right) E_{\theta^0}\left[E^2_{\theta^0}\left[V_1^2|Y\right]\right],$$

which is by applying the *generalised binomial formula* and (4.42) bounded from above by $\frac{M}{K_1}$ times a constant. Altogether I have

$$M^2\,|\phi_1| \leqslant \frac{M^3}{K_1}C,$$

where C is a constant. By definition of M in (4.38) it holds that

$$M^3 \leqslant K_1^{\frac{3}{4}},$$

which implies

$$M^2\,|\phi_1| \leqslant \frac{M^3}{K_1}C \leqslant \frac{K_1^{\frac{3}{4}}}{K_1}C = K_1^{-\frac{1}{4}}C.$$

By assumption on K_1 it holds that $K_1^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2\phi_1 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.3.2 Case 1: Convergence of ϕ_2

For ϕ_2 in case 1 I get by (4.41) and application of Lemma 4.9 the expression

$$\begin{aligned}
M^2\phi_2 &= \frac{M^2}{D}\left(\frac{D}{{\sigma_V^2}^2} - \frac{3}{{\sigma_V^2}^4}E_{\theta^2}\left[V^TV|Y\right] + \frac{1}{{\sigma_V^2}^6}Var_{\theta^2}\left(V^TV|Y\right)\right.\\
&\quad \left. -E_{\theta^2}\left[\frac{D}{{\sigma_V^2}^2} - \frac{3}{{\sigma_V^2}^4}E_{\theta^2}\left[V^TV|Y\right] + \frac{1}{{\sigma_V^2}^6}Var_{\theta^2}\left(V^TV|Y\right)\right]\right)\\
&= \frac{M^2}{{\sigma_V^2}^2} - \frac{3M^2}{D{\sigma_V^2}^4}\sum_{i=1}^{D}E_{\theta^2}\left[V_i^2|Y\right] + \frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D}E^2_{\theta^2}\left[V_1^4|Y\right]\\
&\quad - \frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D}E^2_{\theta^2}\left[V_i^2|Y\right] - \frac{M^2}{{\sigma_V^2}^2} + \frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D}E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right]\\
&= -\frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D}\left(3{\sigma_V^2}^2E_{\theta^2}\left[V_i^2|Y\right] - E_{\theta^2}\left[V_i^4|Y\right]\right)\\
&\quad - \frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D}\left(E^2_{\theta^2}\left[V_i^2|Y\right] - E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right]\right).
\end{aligned}$$

By the *Law of Large Numbers* $\frac{1}{D}\sum_{i=1}^{D}E_{\theta^2}\left[V_i^2|Y\right]$ and $\frac{1}{D}\sum_{i=1}^{D}E_{\theta^2}\left[V_i^4|Y\right]$ converge with rate D as $n \to \infty$ to $E_{\theta^2}\left[V_1^2\right] = {\sigma_V^2}^2$ and $E_{\theta^2}\left[V_1^4\right] = 3{\sigma_V^2}^4$, respectively. Thus it holds for the first summand of $M^2\phi_2$ that

$$\frac{M^2}{{\sigma_V^2}^6}\sum_{i=1}^{D}\left(3{\sigma_V^2}^2E_{\theta^2}\left[V_i^2|Y\right] - E_{\theta^2}\left[V_i^4|Y\right]\right) \in M^2o\left(D\right),$$

where $M^2 \leqslant K_1^{\frac{1}{2}} = D^{\frac{1}{2}}$ by definition of M – see (4.38). Hence

$$\frac{M^2}{{\sigma_V^2}^6}\sum_{i=1}^{D}\left(3{\sigma_V^2}^2E_{\theta^2}\left[V_i^2|Y\right] - E_{\theta^2}\left[V_i^4|Y\right]\right) \in o\left(D^{\frac{1}{2}}\right).$$

For this reason the first summand is $\frac{1}{D}o\left(D^{\frac{1}{2}}\right) = o(D^{-\frac{1}{2}})$ and thus converges to zero. Since all V_i are independently and identically distributed, so is the projection of their transformations on the measure space of Y, which for their expectations leads to

$$\frac{1}{D}\sum_{i=1}^{D}E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right] = E_{\theta^2}\left[E^2_{\theta^2}\left[V_1^2|Y\right]\right].$$

By the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variable converges to its expectation and thus for the random variables $E^2_{\theta^2}\left[V_i^2|Y\right]$ I get

$$\frac{1}{D}\sum_{i=1}^{D} E^2_{\theta^2}\left[V_i^2|Y\right] \to E_{\theta^2}\left[E^2_{\theta^2}\left[V_1^2|Y\right]\right]$$

with rate D, which is why I have for the second summand

$$\frac{M^2}{{\sigma_V^2}^6}\left(\sum_{i=1}^{D} E^2_{\theta^2}\left[V_i^2|Y\right] - E_{\theta^2}\left[E^2_{\theta^2}\left[V_1^2|Y\right]\right]\right) \in M^2 o\left(D\right) = o\left(D^{\frac{1}{2}}\right).$$

Thus the second summand is $\frac{1}{D}o\left(D^{\frac{1}{2}}\right) = o(D^{-\frac{1}{2}})$ and thus also converges to zero as $n \to \infty$ (and thus $D \to \infty$), which leads altogether to

$$M^2\phi_2 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.3.3 Case 1: Convergence of ϕ_3

For ϕ_3 in case 1 I get by (4.40) and (4.41) the expression

$$\begin{aligned}
M^2\phi_3 &= \frac{M^2}{D}\Bigg(E_{\theta^2}\left[\frac{D}{{\sigma_V^2}^2} - \frac{3}{{\sigma_V^2}^4}E_{\theta^2}\left[V^TV|Y\right] + \frac{1}{{\sigma_V^2}^6}Var_{\theta^2}\left(V^TV|Y\right)\right] \\
&\qquad - E_{\theta^2}\left[\frac{D}{{\sigma_V^0}^2} - \frac{3}{{\sigma_V^0}^4}E_{\theta^0}\left[V^TV|Y\right] + \frac{1}{{\sigma_V^0}^6}Var_{\theta^0}\left(V^TV|Y\right)\right]\Bigg) \\
&= \frac{M^2}{{\sigma_V^2}^2} - \frac{M^2}{D{\sigma_V^2}^6}\sum_{i=1}^{D} E_{\theta^2}\left[E^2_{\theta^2}\left[V_i1^2|Y\right]\right] - \frac{M^2}{{\sigma_V^0}^2} \\
&\quad + \frac{3M^2}{D{\sigma_V^0}^4}\sum_{i=1}^{D} E_{\theta^2}\left[E_{\theta^0}\left[V_i^2|Y\right]\right] - \frac{M^2}{D{\sigma_V^0}^6}\sum_{i=1}^{D} E_{\theta^2}\left[E_{\theta^0}\left[V_i^4|Y\right]\right] \\
&\quad + \frac{M^2}{D{\sigma_V^0}^6}\sum_{i=1}^{D} E_{\theta^2}\left[E^2_{\theta^0}\left[V_i^2|Y\right]\right] \\
&= M^2\left(\frac{1}{{\sigma_V^2}^2} - \frac{1}{{\sigma_V^0}^2}\right) \\
&\quad - \frac{M^2}{D}\sum_{i=1}^{D}\left(\frac{1}{{\sigma_V^2}^6}E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right] - \frac{1}{{\sigma_V^0}^6}E_{\theta^2}\left[E^2_{\theta^0}\left[V_i^2|Y\right]\right]\right) \\
&\quad + \frac{M^2}{D{\sigma_V^0}^6}\sum_{i=1}^{D}\left(3{\sigma_V^0}^2 E_{\theta^2}\left[E_{\theta^0}\left[V_i^2|Y\right]\right] - E_{\theta^2}\left[E_{\theta^0}\left[V_i^4|Y\right]\right]\right).
\end{aligned}$$

As shown in Section 4.2.2.3.1 the first summand converges to zero as $n \to \infty$. By the definition of $N_n(\theta^0)$ the random variable $E^2_{\theta^2}\left[V_i^2|Y\right]$

converges as $n \to \infty$ to $E^2_{\theta^0}\left[V_i^2|Y\right]$. For this reason and V_i being iid the second summand of ϕ_3 converges by the *Law of Large Numbers* to

$$\frac{1}{D}\sum_{i=1}^{D}\left(\frac{1}{{\sigma_V^2}^6}E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right]-\frac{1}{{\sigma_V^0}^6}E_{\theta^2}\left[E^2_{\theta^0}\left[V_i^2|Y\right]\right]\right)$$
$$\to\left(\frac{1}{{\sigma_V^2}^6}-\frac{1}{{\sigma_V^0}^6}\right)E_{\theta^2}\left[E^2_{\theta^0}\left[V_1^2|Y\right]\right],$$

which is by applying the *generalised binomial formula* and (4.42) bounded from above by $\frac{M}{K_1}$ times a constant. Therefore

$$\frac{M^2}{D}\sum_{i=1}^{D}\left(\frac{1}{{\sigma_V^2}^6}E_{\theta^2}\left[E^2_{\theta^2}\left[V_i^2|Y\right]\right]-\frac{1}{{\sigma_V^0}^6}E_{\theta^2}\left[E^2_{\theta^0}\left[V_i^2|Y\right]\right]\right)$$

is bounded from above by $\frac{M^3}{K_1}C \leqslant K^{-\frac{1}{4}}C$, where C is a constant and which converges to zero as $n \to \infty$. By the *Law of Large Numbers* $\frac{1}{D}\sum_{i=1}^{D}E_{\theta^0}\left[V_i^2|Y\right]$ and $\frac{1}{n}\sum_{i=1}^{D}E_{\theta^0}\left[V_i^4|Y\right]$ converge with rate D as $n \to \infty$ to $E_{\theta^0}\left[V_1^2\right]={\sigma_V^0}^2$ and $E_{\theta^0}\left[V_1^4\right]=3{\sigma_V^0}^4$, respectively. This leads for the expectations in the third summand to

$$E_{\theta^2}\left[E_{\theta^0}\left[V_i^2|Y\right]\right]\to E_{\theta^2}\left[{\sigma_V^0}^2\right]={\sigma_V^0}^2$$
and
$$E_{\theta^2}\left[E_{\theta^0}\left[V_i^4|Y\right]\right]\to E_{\theta^2}\left[3{\sigma_V^0}^4\right]=3{\sigma_V^0}^4.$$

Hence for the third summand I have by

$$\frac{M^2}{{\sigma_V^0}^6}\sum_{i=1}^{D}\left(3{\sigma_V^0}^2E_{\theta^2}\left[E_{\theta^0}\left[V_i^2|Y\right]\right]-E_{\theta^2}\left[E_{\theta^0}\left[V_i^4|Y\right]\right]\right)$$
$$\in M^2 o(D)=o\left(D^{\frac{1}{2}}\right),$$

Thus this summand is $\frac{1}{D}o\left(D^{\frac{1}{2}}\right)=o(D^{-\frac{1}{2}})$ and thus it converges to zero as $n \to \infty$. As a result I have

$$M^2\phi_3 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.3.4 Case 1: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and I can finally conclude

$$M^2\phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.3.5 Case 1: Convergence of ϕ_5

For ϕ_5 I have in all cases

$$|\phi_5| \leqslant \kappa$$

with κ as defined in (4.37) and hence with the definition of M – see (4.38) –

$$M^2 |\phi_5| \leqslant M^2 \kappa \leqslant \kappa^{-\frac{1}{2}} \kappa = \kappa^{\frac{1}{2}} \to 0$$

by Assumption 1, which I proved in Section 4.2.1.

4.2.2.4 *Case 2*

Employing Lemma 4.10 the expectation (4.39) in this case can be expressed by

$$\begin{aligned} &\frac{1}{\sqrt{D}\sqrt{n}} E_\theta \left[\frac{\partial^2}{\partial \sigma_V \partial \sigma} \ell(\theta|Y) \Big|_{\theta^\star} \right] \\ &\quad = \frac{1}{\sqrt{D}\sqrt{n} {\sigma^\star}^2 {\sigma_V^\star}^3} E_\theta \left[\mathrm{Cov}_{\theta^\star} \left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n \right)^T \varepsilon, V^T V | Y \right) \right]. \quad (4.43) \end{aligned}$$

Whenever $\theta = \theta^\star$ this expression simplifies with condition (B6) and thus $E[V_i|Y]$ and $E[\varepsilon_{ij}|Y]$ pairwise independent from each other. I have

$$\begin{aligned} &\frac{1}{\sqrt{D}\sqrt{n}} E_{\theta^\star} \left[\frac{\partial^2}{\partial \sigma_V \partial \sigma} \ell(\theta|Y) \Big|_{\theta^\star} \right] \\ &\quad = \frac{1}{\sqrt{D}\sqrt{n} {\sigma^\star}^2 {\sigma_V^\star}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \Big(E_{\theta^\star} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} V_i^2 \right] \\ &\qquad - E_{\theta^\star} \left[E_{\theta^\star} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] E_{\theta^\star} \left[V_i^2 | Y \right] \right] \Big) \\ &\quad = 0. \quad (4.44) \end{aligned}$$

4.2.2.4.1 Case 2: Convergence of ϕ_1

For ϕ_1 in case 2 I have the expression

$$\begin{aligned} &M^2 \phi_1 \\ &\quad = \frac{M^2}{\sqrt{D}\sqrt{n}} \sup_{\theta^1 \in N_n(\theta^0)} \Bigg(\frac{1}{{\sigma^1}^2 {\sigma_V^1}^3} \mathrm{Cov}_{\theta^1} \left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n \right)^T \varepsilon, V^T V | Y \right) \\ &\qquad - \frac{1}{{\sigma^2}^2 {\sigma_V^2}^3} \mathrm{Cov}_{\theta^2} \left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n \right)^T \varepsilon, V^T V | Y \right) \Bigg), \end{aligned}$$

where, as $n \to \infty$, both covariances converge to zero with rate $\sqrt{D}\sqrt{n}$ with the same argumentation as in the proof of Lemma 4.15 of Case 2. Altogether I have

$$M^2\phi_1 \in \frac{M^2}{\sqrt{D}\sqrt{n}} o\left(\sqrt{D}\sqrt{n}\right),$$

where $M^2 \leqslant D^{-\frac{1}{4}}n^{-\frac{1}{4}}$ by (4.38) and thus

$$M^2\phi_1 \in o\left(D^{-\frac{1}{4}}n^{-\frac{1}{4}}\right).$$

4.2.2.4.2 Case 2: Convergence of ϕ_2

For ϕ_2 in case 2 I get by (4.44) the expression

$$M^2\phi_2 = \frac{M^2}{\sqrt{D}\sqrt{n}{\sigma^2}^2{\sigma_V^2}^3}\left(\mathrm{Cov}_{\theta^2}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T \varepsilon, V^TV|Y\right) - 0\right),$$

where, as $n \to \infty$, the covariance converges to zero with rate $\sqrt{D}\sqrt{n}$ with the same argumentation as in the proof of Lemma 4.15 Case 2. Thus

$$M^2\phi_2 \in \frac{M^2}{\sqrt{D}\sqrt{n}} o\left(\sqrt{D}\sqrt{n}\right) = o\left(D^{-\frac{1}{4}}n^{-\frac{1}{4}}\right).$$

4.2.2.4.3 Case 2: Convergence of ϕ_3

For ϕ_3 in case 2 I get by (4.43) and (4.44) the expression

$$M^2\phi_3 = -\frac{M^2}{\sqrt{D}\sqrt{n}{\sigma^0}^2{\sigma_V^0}^3}E_{\theta^2}\left[\mathrm{Cov}_{\theta^0}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T \varepsilon, V^TV|Y\right)\right],$$

where $\frac{1}{\sqrt{D}\sqrt{n}}\mathrm{Cov}_{\theta^0}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T \varepsilon, V^TV|Y\right)$ converges with rate $\sqrt{D}\sqrt{n}$ to zero as $n \to \infty$. The application of the *Lebesgue Theorem* allows me to interchange expectation and limes. I get

$$M^2|\phi_3| \in \frac{M^2}{\sqrt{D}\sqrt{n}} o\left(\sqrt{D}\sqrt{n}\right) = o\left(D^{-\frac{1}{4}}n^{-\frac{1}{4}}\right).$$

4.2.2.4.4 Case 2: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and can finally conclude

$$M^2\phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.4.5 Case 2: Convergence of ϕ_5

For ϕ_5 see Section 4.2.2.3.5.

4.2.2.5 *Case 3*

Employing Lemma 4.11 the expectation (4.39) in this case can be expressed by

$$\frac{1}{\sqrt{D}\sqrt{n}} E_\theta \left[\frac{\partial^2}{\partial \sigma_V \partial \beta_\tau} \ell(\theta|Y) \bigg|_{\theta^\star} \right] = -\frac{1}{\sqrt{D}\sqrt{n}\sigma^\star {\sigma_V^\star}^3} E_\theta \left[\mathrm{Cov}_{\theta^\star} \left(V^T V, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right]. \tag{4.45}$$

Whenever $\theta = \theta^\star$ this expression simplifies with condition (B6). Thus the conditional expectations of transformations of $V_i|Y$ and $\varepsilon_{ij}|Y$ pairwise independent from each other and I have

$$\begin{aligned} &\frac{1}{\sqrt{D}\sqrt{n}} E_{\theta^\star} \left[\frac{\partial^2}{\partial \sigma_V \partial \beta_{\tau,h}} \ell(\theta|Y) \bigg|_{\theta^\star} \right] \\ &= \frac{1}{\sqrt{D}\sqrt{n}\sigma^\star {\sigma_V^\star}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \Big(E_{\theta^\star} \left[V_i^2 \mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] x_{ij} \\ &\qquad - E_{\theta^\star} \left[E_{\theta^\star} \left[V_i^2 | Y \right] \right] E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} \Big) \\ &= 0_p. \end{aligned} \tag{4.46}$$

4.2.2.5.1 Case 3: Convergence of ϕ_1

For ϕ_1 in case 3 I have the expression

$$\begin{aligned} &M^2 \phi_1 \\ &= \frac{M^2}{\sqrt{D}\sqrt{n}} \sup_{\theta^1 \in N_n(\theta^0)} \left(\frac{1}{{\sigma^1}^2 {\sigma_V^1}^3} \mathrm{Cov}_{\theta^1} \left(V^T V, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right. \\ &\qquad \left. - \frac{1}{{\sigma^2}^2 {\sigma_V^2}^3} \mathrm{Cov}_{\theta^2} \left(V^T V, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right), \end{aligned}$$

where, as $n \to \infty$, both covariances converge to zero with rate $\sqrt{D}\sqrt{n}$ with the same argumentation as in the proof of Lemma 4.15 Case 3. Altogether I have

$$M^2 \phi_1 \in \frac{M^2}{\sqrt{D}\sqrt{n}} o\left(\sqrt{D}\sqrt{n}\right),$$

where $M^2 \leqslant D^{-\frac{1}{4}} n^{-\frac{1}{4}}$ by (4.38) and thus

$$M^2 \phi_1 \in o\left(D^{-\frac{1}{4}} n^{-\frac{1}{4}}\right).$$

4.2.2.5.2 Case 3: Convergence of ϕ_2

For ϕ_2 in case 3 I get by (4.46) the expression

$$M^2\phi_2 = \frac{M^2}{\sqrt{D}\sqrt{n}{\sigma^2}^2{\sigma_V^2}^3}\left(\mathrm{Cov}_{\theta^2}\left(V^TV, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X} - 0\right),$$

where, as $n \to \infty$, the covariance converges to zero with rate $\sqrt{D}\sqrt{n}$ with the same argumentation as in the proof of Lemma 4.15 Case 3. Thus

$$M^2\phi_2 \in \frac{M^2}{\sqrt{D}\sqrt{n}}o\left(\sqrt{D}\sqrt{n}\right) = o\left(D^{-\frac{1}{4}}n^{-\frac{1}{4}}\right).$$

4.2.2.5.3 Case 3: Convergence of ϕ_3

For ϕ_3 in case 3 I get by (4.45) and (4.46) the expression

$$M^2\phi_3 = -\frac{M^2}{\sqrt{D}\sqrt{n}{\sigma^0}^2{\sigma_V^0}^3}E_{\theta^2}\left[\mathrm{Cov}_{\theta^0}\left(V^TV, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}\right],$$

where $\frac{1}{\sqrt{D}\sqrt{n}}\mathrm{Cov}_{\theta^0}\left(V^TV, \mathbf{1}^T_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}$ converges with rate $\sqrt{D}\sqrt{n}$ to zero as $n \to \infty$. Applying the *Lebesgue Theorem* allows me to interchange expectation and limes I get

$$M^2\left|\phi_3\right| \in \frac{M^2}{\sqrt{D}\sqrt{n}}o\left(\sqrt{D}\sqrt{n}\right) = o\left(D^{-\frac{1}{4}}n^{-\frac{1}{4}}\right).$$

4.2.2.5.4 Case 3: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and can finally conclude

$$M^2\phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.5.5 Case 3: Convergence of ϕ_5

For ϕ_5 see Section 4.2.2.3.5.

4.2.2.6 *Case 4*

The expectation (4.39) in this case can be expressed by

$$\begin{aligned}
&\frac{1}{n}E_\theta\left[\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)\Big|_{\theta^\star}\right] \\
&\quad = \frac{1}{n}E_\theta\left[\frac{n}{{\sigma^\star}^2} - \frac{2}{{\sigma^\star}^3}\left(E_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}_n^T E_{\theta^\star}\left[\varepsilon|Y\right]\right)\right. \\
&\qquad \left. + \frac{1}{{\sigma^\star}^4}\mathrm{Var}_{\theta^\star}\left(\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right)\right].
\end{aligned}$$

The variance may be split as follows

$$\begin{aligned}&\frac{1}{n}E_\theta\left[Var_{\theta^\star}\left(\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right)\right]\\&\quad=\frac{1}{n}\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]-E_\theta\left[E^2_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\right),\end{aligned}$$

which leads to

$$\begin{aligned}&\frac{1}{n}E_\theta\left[\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)\bigg|_{\theta^\star}\right]\\&\quad=\frac{1}{\sigma^{\star 2}}-\frac{2}{n\sigma^{\star 3}}\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]-\tau\mathbf{1}_n^TE_\theta\left[E_{\theta^\star}\left[\varepsilon|Y\right]\right]\right)\\&\qquad+\frac{1}{n\sigma^{\star 4}}\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]-E_\theta\left[E^2_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\right)\\&\quad=\frac{1}{\sigma^{\star 2}}-\frac{2}{n\sigma^{\star 3}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]\right]\\&\qquad+\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_\theta\left[E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon^2_{ij}|Y\right]\right]\\&\qquad-\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_\theta\left[E^2_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right].\end{aligned}\tag{4.47}$$

Whenever $\theta=\theta^\star$ this expression simplifies with $\varepsilon_{ij}\overset{iid}{\sim}ALD(0,\sigma^\star,\tau)$, $E_{\theta^\star}\left[E_{\theta^\star}\left[\cdot|Y\right]\right]=E_{\theta^\star}\left[\cdot\right]$, and Corollary 2.2 to

$$\begin{aligned}&\frac{1}{n}E_{\theta^\star}\left[\frac{\partial^2}{\partial\sigma^2}\ell(\theta|Y)\bigg|_{\theta^\star}\right]\\&\quad=\frac{1}{\sigma^{\star 2}}-\frac{2}{n\sigma^{\star 3}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}\right]\\&\qquad+\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon^2_{ij}\right]\\&\qquad-\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_{\theta^\star}\left[E^2_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right]\\&\quad=\frac{1}{\sigma^{\star 2}}-\frac{2}{\sigma^{\star 2}}+\frac{2\tau}{\sigma^{\star 2}(1-\tau)^2}\\&\qquad-\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_{\theta^\star}\left[E^2_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right]\\&\quad=-\frac{1}{\sigma^{\star 2}}+\frac{2\tau}{\sigma^{\star 2}(1-\tau)^2}-\frac{1}{n\sigma^{\star 4}}\sum_{i=1}^D\sum_{j=1}^{n_i}E_{\theta^\star}\left[E^2_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right].\end{aligned}\tag{4.48}$$

4.2.2.6.1 Case 4: Convergence of ϕ_1

For ϕ_1 in case 4 I have the expression

$$\begin{aligned}
M^2\phi_1 = M^2 \sup_{\theta^1 \in N_n(\theta^0)} \Bigg(& -\frac{1}{{\sigma^1}^2} \\
& + \frac{2}{n{\sigma^1}^3}\left(E_{\theta^1}\left[\mathbf{1}^T_{\{\varepsilon \leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}^T_n E_{\theta^1}\left[\varepsilon|Y\right]\right) \\
& + \frac{1}{n{\sigma^1}^4} Var_{\theta^1}\left(\mathbf{1}^T_{\{\varepsilon \leqslant 0\}}\varepsilon|Y\right) + \frac{1}{{\sigma^2}^2} \\
& - \frac{2}{n{\sigma^2}^3}\left(E_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon \leqslant 0\}}\varepsilon|Y\right] - \tau\mathbf{1}^T_n E_{\theta^2}\left[\varepsilon|Y\right]\right) \\
& - \frac{1}{n{\sigma^2}^4} Var_{\theta^2}\left(\mathbf{1}^T_{\{\varepsilon \leqslant 0\}}\varepsilon|Y\right)\Bigg) \\
= M^2 \sup_{\theta^1 \in N_n(\theta^0)} \Bigg(& \frac{1}{{\sigma^2}^2} - \frac{1}{{\sigma^1}^2} \\
& + \frac{2}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\Bigg(\frac{1}{{\sigma^1}^3}E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] \\
& \qquad - \frac{1}{{\sigma^2}^3}E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right]\Bigg) \\
& + \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\Bigg(\frac{1}{{\sigma^1}^4}E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}^2|Y\right] \\
& \qquad - \frac{1}{{\sigma^2}^4}E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}^2|Y\right]\Bigg) \\
& - \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\Bigg(\frac{1}{{\sigma^1}^4}E^2_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}|Y\right] \\
& \qquad - \frac{1}{{\sigma^2}^4}E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}|Y\right]\Bigg)\Bigg),
\end{aligned}$$

where each summand converges to zero as $n \to \infty$ as shown in the following: Following the argumentation in Section 4.2.2.3.1 I have for the first summand the inequality

$$\sup_{\sigma^1_V \in N_{2,n}(\theta^0)} \left|\sigma^1 - \sigma^2\right| \leqslant 2\frac{M}{K_2}, \tag{4.49}$$

where $N_{2,n}$ is the projection on the second dimension of the $(p+2)$-dimensional set N_n. Thus I get by (4.49) the boundedness

$$\sup_{\sigma^1_V \in N_{2,n}(\theta^0)} \left|\frac{1}{{\sigma^2}^2} - \frac{1}{{\sigma^1}^2}\right| \leqslant \frac{2M}{K_2{\sigma^1}^2{\sigma^2}^2}.$$

By the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]$ and $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]$ converge as $n\to\infty$ to $E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}-\tau\right)\varepsilon_{1,1}\right]=\sigma^1$ and $E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}-\tau\right)\varepsilon_{1,1}\right]=\sigma^2$, respectively. Hence the second summand converges to

$$\frac{1}{n}\left(\frac{2E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]}{{\sigma^1}^3}-\frac{2E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]}{{\sigma^2}^3}\right)$$
$$\to\frac{2\sigma^1}{{\sigma^1}^3}-\frac{2\sigma^2}{{\sigma^2}^3}=2\left(\frac{1}{{\sigma^1}^2}-\frac{1}{{\sigma^2}^2}\right),$$

which is by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. In the third summand I can apply the *Law of Large Numbers*, again. It holds that $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}^2|Y\right]$ and $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}^2|Y\right]$ converge as $n\to\infty$ to $E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}^2\right]=\frac{2\tau{\sigma^1}^2}{(1-\tau)^2}$ and $E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}^2\right]=\frac{2\tau{\sigma^2}^2}{(1-\tau)^2}$, respectively. Thus

$$\frac{1}{n}\left(\frac{E_{\theta^1}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}_{\{\varepsilon\leqslant 0\}}{}^T\varepsilon|Y\right]}{{\sigma^1}^4}.-\frac{E_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]}{{\sigma^2}^4}\right)$$
$$\to\frac{2\tau}{{\sigma^1}^2(1-\tau)^2}-\frac{2\tau}{{\sigma^2}^2(1-\tau)^2}=\frac{2\tau}{(1-\tau)^2}\left(\frac{1}{{\sigma^1}^2}-\frac{1}{{\sigma^2}^2}\right),$$

which is by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. By the definition of $N_n(\theta^0)$ the random variables $E^2_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ and $E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ converge as $n\to\infty$ to $E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$. At the same time by the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variables converges to its expectation and thus for the random variables $E^2_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ and $E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ I get as $n\to\infty$

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E^2_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\to E_{\theta^0}\left[E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]$$

and

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\to E_{\theta^0}\left[E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right],$$

where $\varepsilon_{1,1}$ is a representative of the p-dimensional error term ε. Hence as $n \to \infty$ I have for the fourth summand

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(\frac{E^2_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]}{{\sigma^1}^4} - \frac{E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]}{{\sigma^2}^4}\right)$$
$$\to \left(\frac{1}{{\sigma^1}^4} - \frac{1}{{\sigma^2}^4}\right) E_{\theta^0}\left[E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right],$$

which is by applying the *generalised binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. Altogether, I have

$$M^2\,|\varphi_1| \leqslant \frac{M^3}{K_2}C,$$

where C is a constant. By definition of M in (4.38) it holds that

$$M^3 \leqslant K_2^{\frac{3}{4}},$$

which implies

$$M^2\,|\varphi_1| \leqslant \frac{M^3}{K_2}C \leqslant \frac{K_2^{\frac{3}{4}}}{K_1}C = K_2^{-\frac{1}{4}}C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2\varphi_1 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.6.2 Case 4: Convergence of φ_2

For φ_2 in case 4 I get by (4.48) and application of Lemma 4.12 the expression

$$\begin{aligned}M^2\varphi_2 = M^2 \Bigg(&\frac{1}{{\sigma^2}^2} + \frac{2}{n{\sigma^2}^3}E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}} - \tau\mathbf{1}_n\right)^T \varepsilon|Y\right]\\ &+\frac{1}{n{\sigma^2}^4}\left(E_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right] - E^2_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right)\\ &-\frac{1}{{\sigma^2}^2} - \frac{2\tau}{{\sigma^2}^2(1-\tau)^2}\\ &+\frac{1}{n{\sigma^2}^4}E_{\theta^2}\left[E^2_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\Bigg)\end{aligned}$$

$$= M^2 \left(\frac{2}{n\sigma^{2^3}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] - \sigma^2 \right) \right.$$
$$+ \frac{1}{n\sigma^{2^4}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij}^2 | Y \right] - \frac{2\tau\sigma^{2^2}}{(1-\tau)^2} \right)$$
$$- \frac{1}{n\sigma^{2^4}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}^2 \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right.$$
$$\left. \left. - \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[E_{\theta^2}^2 \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] \right) \right) .$$

By the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right]$ and $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij}^2 | Y \right]$ converge with rate n as $n \to \infty$ to $E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}} - \tau \right) \varepsilon_{1,1} \right] = \sigma^2$ and $E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}} - \tau \right) \varepsilon_{1,1}^2 \right] = \frac{2\tau\sigma^{2^2}}{(1-\tau)^2}$, respectively. Thus it holds for the first and the second summand of $M^2\phi_2$ that

$$\frac{2M^2}{\sigma^{2^3}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] - \sigma^2 \right) \in M^2 o(n)$$

and

$$\frac{M^2}{\sigma^{2^4}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij}^2 | Y \right] - \frac{2\tau\sigma^{2^2}}{(1-\tau)^2} \right) \in M^2 o(n),$$

where $M^2 \leqslant K_2^{\frac{1}{2}} = n^{\frac{1}{2}}$ by definition of M – see (4.38). Hence

$$\frac{2M^2}{n\sigma^{2^3}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] - \sigma^2 \right) \in o\left(n^{-\frac{1}{2}}\right)$$

and

$$\frac{M^2}{n\sigma^{2^4}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij}^2 | Y \right] - \frac{2\tau\sigma^{2^2}}{(1-\tau)^2} \right) \in o\left(n^{-\frac{1}{2}}\right).$$

For this reason the first and the second summand converge to zero. Since all ε_{ij} are independently and identically distributed, so is the projection of their transformations on the measure space of Y, which for their expectations leads to

$$\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[E_{\theta^2}^2 \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] = E_{\theta^2} \left[E_{\theta^2}^2 \left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}} \varepsilon_{1,1} | Y \right] \right].$$

By the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variable converges to its expectation. Thus for the random variables $E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ I get

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\to E_{\theta^2}\left[E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}|Y\right]\right]$$

with rate n, which is why I have for the third summand

$$\frac{M^2}{n{\sigma^2}^4}\left(\sum_{i=1}^{D}\sum_{j=1}^{n_i}E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]-\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^2}\left[E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right]\right)$$
$$\in\frac{M^2}{n}o\left(n\right)=o\left(n^{-\frac{1}{2}}\right).$$

Thus this summand also converges to zero as $n\to\infty$, which leads altogether to

$$M^2\phi_2\to 0\quad\text{as } n\to\infty.$$

4.2.2.6.3 Case 4: Convergence of ϕ_3

For ϕ_3 in case 4 I get by (4.47) and (4.48)

$$\begin{aligned}
&M^2\phi_3\\
&\quad=M^2\left(-\frac{1}{{\sigma^2}^2}+\frac{2\tau}{{\sigma^2}^2(1-\tau)^2}\right.\\
&\qquad-\frac{1}{n{\sigma^2}^4}E_{\theta^2}\left[E^2_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\\
&\qquad-\frac{1}{{\sigma^0}^2}-\frac{2}{n{\sigma^0}^3}E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]\right]\\
&\qquad-\frac{1}{n{\sigma^0}^4}E_{\theta^2}\left[E_{\theta^0}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\\
&\qquad\left.+\frac{1}{n{\sigma^0}^4}E_{\theta^2}\left[E^2_{\theta^0}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]\right]\right)
\end{aligned}$$

$$
\begin{aligned}
&= M^2 \left(\frac{1}{\sigma^{2^2}} - \frac{1}{\sigma^{0^2}} \right.\\
&\quad + \left(\frac{2}{\sigma^{2^2}} - \frac{2}{n\sigma^{0^3}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \right] \right)\\
&\quad + \left(\frac{2\tau}{\sigma^{2^2}(1-\tau)^2} - \frac{1}{n\sigma^{0^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij}^2 | Y \right] \right] \right)\\
&\quad - \left(\frac{1}{n\sigma^{2^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^2}^2\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] \right.\\
&\quad \left.\left. - \frac{1}{n\sigma^{0^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}^2\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] \right) \right).
\end{aligned}
$$

As shown in Section 4.2.2.6.1 – cf. (4.49) – the first summand is bounded from above by $\frac{M}{K_1}$ times a constant. By the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right]$ and $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}^2|Y\right]$ converge as $n \to \infty$ to $E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}} - \tau\right)\varepsilon_{1,1}\right] = \sigma^0$ and $E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\varepsilon_{1,1}^2\right] = \frac{2\tau\sigma^{0^2}}{(1-\tau)^2}$, respectively. For the second and the third summand of ϕ_3 this leads to

$$
\begin{aligned}
&\frac{2}{\sigma^{2^2}} - \frac{2}{n\sigma^{0^3}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \mathbf{1}_n \right) \varepsilon_{ij} | Y \right] \right]\\
&\quad \to \frac{2}{\sigma^{2^2}} - \frac{2}{n\sigma^{0^3}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[\sigma^0\right] = \frac{2}{\sigma^{2^2}} - \frac{2}{\sigma^{0^2}}
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{2\tau}{\sigma^{2^2}(1-\tau)^2} - \frac{1}{n\sigma^{0^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij}^2 | Y \right] \right]\\
&\quad \to \frac{2\tau}{\sigma^{2^2}(1-\tau)^2} - \frac{1}{n\sigma^{0^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[\frac{2\tau\sigma^{0^2}}{(1-\tau)^2} \right]\\
&\quad = \frac{2\tau}{\sigma^{2^2}(1-\tau)^2} - \frac{2\tau}{\sigma^{0^2}(1-\tau)^2} = \frac{2\tau}{(1-\tau)^2}\left(\frac{1}{\sigma^{2^2}} - \frac{1}{\sigma^{0^2}} \right),
\end{aligned}
$$

which are both by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. By construction of $N_n(\theta^0)$ the random variable $E_{\theta^2}^2\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$

converges as $n \to \infty$ to $E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}|Y\right]$. For this reason and ε_{ij} being iid the last summand of ϕ_3 converges to

$$\begin{aligned}
&\frac{1}{n\sigma^{2^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E^2_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}|Y\right]\right] \\
&- \frac{1}{n\sigma^{0^4}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2}\left[E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}}\varepsilon_{ij}|Y\right]\right] \\
&\to \left(\frac{1}{\sigma^{2^4}} - \frac{1}{\sigma^{0^4}}\right) E_{\theta^2}\left[E^2_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{1,1} \leqslant 0\}}\varepsilon_{1,1}|Y\right]\right],
\end{aligned}$$

which is by applying the *generalised binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant C. Altogether I have

$$M^2 |\phi_3| \leqslant \frac{M^3}{K_2} C \leqslant \frac{K_2^{\frac{3}{4}}}{K_2} C = K_2^{-\frac{1}{4}} C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2 \phi_3 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.6.4 Case 4: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and can finally conclude

$$M^2 \phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.6.5 Case 4: Convergence of ϕ_5

For ϕ_5 see Section 4.2.2.3.5.

4.2.2.7 *Case 5*

Employing Lemma 4.13 the expectation (4.39) in this case can be expressed by

$$\begin{aligned}
&\frac{1}{n} E_\theta\left[\frac{\partial^2}{\partial\sigma\partial\beta_\tau} \ell(\theta|Y)\bigg|_{\theta^\star}\right] \\
&= \frac{1}{n} E_\theta\left[-\frac{1}{\sigma^{\star 2}}\left(E_{\theta^\star}\left[\mathbf{1}^T_{\{\varepsilon \leqslant 0\}}|Y\right] - \tau \mathbf{1}_n^T\right)\mathbf{X}\right. \\
&\qquad\left. - \frac{1}{\sigma^{\star 3}} Cov_{\theta^\star}\left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n\right)^T \varepsilon, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}}|Y\right)\mathbf{X}\right].
\end{aligned}$$

The covariance may be split as follows

$$
\begin{aligned}
&\frac{1}{n}E_\theta\left[\mathrm{Cov}_{\theta^\star}\left(\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon,\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right)\right]\\
&\quad=\frac{1}{n}\left(E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right.\\
&\qquad\left.-E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right),
\end{aligned}
$$

which leads to

$$
\begin{aligned}
&\frac{1}{n}E_\theta\left[\frac{\partial^2}{\partial\sigma\partial\beta_{\tau,h}}\ell(\theta|Y)\bigg|_{\theta^\star}\right]\\
&=-\frac{1}{n\sigma^{\star 2}}\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]-\tau\mathbf{1}_n^T\right)\mathbf{X}\\
&\quad-\frac{1}{n\sigma^{\star 3}}\left(E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right.\\
&\qquad\left.-E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}}-\tau\mathbf{1}_n\right)^T\varepsilon|Y\right]E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right)\mathbf{X}\\
&=-\frac{1}{n\sigma^{\star 2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]-\tau\right)x_{ij}\\
&\quad-\frac{1}{n\sigma^{\star 3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_\theta\left[E_{\theta^\star}\left[(1-\tau)\,\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right]x_{ij}\\
&\quad+\frac{1}{n\sigma^{\star 3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_\theta\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}.
\end{aligned}
\tag{4.50}
$$

Whenever $\theta=\theta^\star$ this expression simplifies with $\varepsilon_{ij}\overset{iid}{\sim}\mathrm{ALD}(0,\sigma^\star,\tau)$, $E_{\theta^\star}\left[E_{\theta^\star}\left[\cdot|Y\right]\right]=E_{\theta^\star}\left[\cdot\right]$, and Corollary 2.2 to

$$
\begin{aligned}
&\frac{1}{n}E_\theta\left[\frac{\partial^2}{\partial\sigma\partial\beta_{\tau,h}}\ell(\theta|Y)\bigg|_{\theta^\star}\right]\\
&=-\frac{1}{n\sigma^{\star 2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\right]-\tau\right)x_{ij}\\
&\quad-\frac{1}{n\sigma^{\star 3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^\star}\left[(1-\tau)\,\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}\right]x_{ij}\\
&\quad+\frac{1}{n\sigma^{\star 3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^\star}\left[E_{\theta^\star}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}
\end{aligned}
$$

$$= \frac{\tau}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij}$$
$$+ \frac{1}{n\sigma^{\star 3}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^\star} \left[E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] E_{\theta^\star} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) |Y \right] \right] x_{ij}. \tag{4.51}$$

4.2.2.7.1 Case 5: Convergence of ϕ_1

For ϕ_1 in case 5 I have the expression

$$\begin{aligned}
M^2 \phi_1 &= \frac{M^2}{n} \sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{\sigma^{1\,2}} \left(E_{\theta^1} \left[\mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right] - \tau \mathbf{1}^T_n \right) \mathbf{X} \right. \\
&\quad - \frac{1}{\sigma^{1\,3}} \mathrm{Cov}_{\theta^1} \left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n \right)^T \varepsilon, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \\
&\quad + \frac{1}{\sigma^{2\,2}} \left(E_{\theta^2} \left[\mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right] - \tau \mathbf{1}^T_n \right) \mathbf{X} \\
&\quad \left. + \frac{1}{\sigma^{2\,3}} \mathrm{Cov}_{\theta^2} \left(\left(\mathbf{1}_{\{\varepsilon \leqslant 0\}} - \tau \mathbf{1}_n \right)^T \varepsilon, \mathbf{1}^T_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right) \\
&= M^2 \sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(\frac{1}{\sigma^{1\,2}} \left(E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} \right. \right. \\
&\quad \left. - \frac{1}{\sigma^{2\,2}} \left(E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} \right) \\
&\quad - \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(\frac{1}{\sigma^{1\,3}} E_{\theta^1} \left[(1-\tau) \mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] x_{ij} \right. \\
&\quad \left. - \frac{1}{\sigma^{2\,3}} E_{\theta^2} \left[(1-\tau) \mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] x_{ij} \right) \\
&\quad + \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(\frac{1}{\sigma^{1\,3}} E_{\theta^1} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \right. \\
&\quad\quad E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} \\
&\quad\quad - \frac{1}{\sigma^{2\,3}} E_{\theta^2} \left[\left(\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \\
&\quad\quad \left. \left. E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} \right) \right),
\end{aligned}$$

where each summand converges to zero as $n \to \infty$ as shown in the following: By the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ and $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converge as $n \to \infty$ to $E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] = \tau$ and $E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] = \tau$, respectively. Hence the first summand converges to zero. Also by the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ and $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ converge as $n \to \infty$ to $E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}\right] = -\frac{\tau\sigma^1}{1-\tau}$ and $E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}\right] = -\frac{\tau\sigma^2}{1-\tau}$, respectively. Hence I have for the second summand of ϕ_1

$$\frac{1-\tau}{n}\left(\frac{E_{\theta^1}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]}{{\sigma^1}^3} - \frac{E_{\theta^2}\left[\mathbf{1}^T_{\{\varepsilon\leqslant 0\}}\varepsilon|Y\right]}{{\sigma^2}^3}\right)$$
$$\to \frac{-\tau\sigma^1}{{\sigma^1}^3} - \frac{-\tau\sigma^2}{{\sigma^2}^3} = -\tau\left(\frac{1}{{\sigma^1}^2} - \frac{1}{{\sigma^2}^2}\right),$$

which is by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. By the definition of $N_n(\theta^0)$ the random variables $E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ and $E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ converge as $n \to \infty$ to $E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$. At the same time by the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variables converges to its expectation and thus I get

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$$
$$\to E_{\theta^0}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]$$

and

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$$
$$\to E_{\theta^0}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right].$$

Hence I have for the third summand of ϕ_1 that

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(\frac{E_{\theta^1}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^1}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]}{{\sigma^1}^3}\right.$$
$$\left. - \frac{E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]}{{\sigma^2}^3}\right)$$
$$\to \left(\frac{1}{{\sigma^1}^3} - \frac{1}{{\sigma^2}^3}\right) E_{\theta^0}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}} - \tau\right)\varepsilon_{ij}|Y\right] E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right],$$

which is by applying the *generalised binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant C. Altogether I have

$$M^2 |\phi_1| \leqslant \frac{M^3}{K_2} C \leqslant \frac{K_2^{\frac{3}{4}}}{K_1} C = K_2^{-\frac{1}{4}} C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2 \phi_1 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.7.2 Case 5: Convergence of ϕ_2

For ϕ_2 in case 5 I get by (4.51) the expression

$$\begin{aligned}
&M^2 \phi_2 \\
&= M^2 \Bigg(-\frac{1}{n{\sigma^2}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} \\
&\quad - \frac{1}{n{\sigma^2}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] x_{ij} \\
&\quad + \frac{1}{n{\sigma^2}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} \\
&\quad - \frac{\tau}{n{\sigma^2}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} \\
&\quad - \frac{1}{n{\sigma^2}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \Big[E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \\
&\qquad\qquad E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \Big] x_{ij} \Bigg) \\
&= M^2 \Bigg(-\frac{1}{n{\sigma^2}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} \\
&\quad - \frac{1}{n{\sigma^2}^3} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] x_{ij} + \tau \sigma^2 x_{ij} \right) \\
&\quad + \frac{1}{n{\sigma^2}^3} \Bigg(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} \\
&\qquad - \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \Big[E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \\
&\qquad\qquad E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \Big] x_{ij} \Bigg) \Bigg).
\end{aligned}$$

By the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ converges with rate n as $n \to \infty$ to $E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}\right] = \tau$. Thus it holds with (B3) for the first summand of $M^2\phi_2$ that

$$\frac{M^2}{n\sigma^{2^2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]-\tau\right)x_{ij} \in \frac{M^2}{n}o\left(n\right) \overset{(\star)}{=} o\left(n^{-\frac{1}{2}}\right),$$

where $(\star)$ holds because $M^2 \leqslant K_2^{\frac{1}{2}} = n^{\frac{1}{2}}$ by definition of M – see (4.38). By the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[(1-\tau)\,\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]$ converges with rate n as $n \to \infty$ to $E_{\theta^2}\left[(1-\tau)\,\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}\right] = -\tau\sigma^2$. Thus it holds for the second summand of $M^2\phi_2$ with (B3) that

$$\begin{aligned}&\frac{M^2}{n\sigma^{2^2}}\sum_{i=1}^{D}\left(\sum_{j=1}^{n_i}E_{\theta^2}\left[(1-\tau)\,\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]x_{ij}+\tau\sigma^2 x_{ij}\right)\\ &\quad\in\frac{M^2}{n}o\left(n\right)=o\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

For this reason the second summand converges to zero. Since all ε_{ij} are independently and identically distributed, so is the projection of their transformations on the measure space of Y, which for their expectations leads to

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]\\ &\quad=E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}-\tau\right)\varepsilon_{1,1}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right].\end{aligned}$$

By the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variable converges to its expectation and thus for the random variables $E_{\theta^2}\left[\left(\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ I get

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^2}\left[\left(\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\\ &\quad\to E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}-\tau\right)\varepsilon_{1,1}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]\end{aligned}$$

with rate n, which is why I have for the third summand

$$\frac{M^2}{n\sigma^{2^3}}\left(\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right] E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right.$$
$$\left.-\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right] E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]\right)$$
$$\in \frac{M^2}{n} o(n) = o\left(n^{-\frac{1}{2}}\right).$$

Thus this summand also converges to zero as $n \to \infty$, which leads altogether to

$$M^2\phi_2 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.7.3 Case 5: Convergence of ϕ_3

For ϕ_3 in case 5 I get by (4.50) and (4.51) the expression

$$M^2\phi_3 = M^2\left(\frac{\tau}{n\sigma^{2^2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}\right.$$
$$+\frac{1}{n\sigma^{2^3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]\right.$$
$$\left. E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right] x_{ij}$$
$$+\frac{1}{n\sigma^{0^2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(E_{\theta^2}\left[E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]-\tau\right)x_{ij}$$
$$+\frac{1}{n\sigma^{0^3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[(1-\tau)\,\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}\varepsilon_{ij}|Y\right]\right]x_{ij}$$
$$-\frac{1}{n\sigma^{0^3}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]\right.$$
$$\left.\left. E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right] x_{ij}\right)$$

$$= M^2 \left(\frac{\tau}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(\frac{1}{{\sigma^2}^2} x_{ij} \right.\right.$$
$$\left. - \frac{1}{{\sigma^0}^3} E_{\theta^2} \left[E_{\theta^0} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] x_{ij} \right)$$
$$+ \frac{1}{n{\sigma^0}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] - \tau \right) x_{ij}$$
$$+ \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(\frac{1}{{\sigma^2}^3} E_{\theta^2} \left[E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \right.\right.$$
$$\left. E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij}$$
$$- \frac{1}{{\sigma^0}^3} E_{\theta^2} \left[E_{\theta^0} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] \right.$$
$$\left.\left.\left. E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij} \right) \right).$$

By the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^0} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right]$ converges as $n \to \infty$ to $E_{\theta^0} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{1,1} \leqslant 0\}} \varepsilon_{1,1} \right] = -\tau\sigma^0$. For the first summand of ϕ_3 this leads with (B3) to

$$\frac{\tau}{n{\sigma^2}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} + \frac{1}{n{\sigma^0}^3} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[E_{\theta^0} \left[(1-\tau) \mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \varepsilon_{ij} | Y \right] \right] x_{ij}$$
$$\to \frac{\tau}{{\sigma^2}^2} c_0 + \frac{1}{{\sigma^0}^3} E_{\theta^2} \left[-\tau\sigma^0 \right] c_0 = -\tau \left(\frac{1}{{\sigma^2}^2} - \frac{1}{{\sigma^0}^2} \right) c_0,$$

where c_0 is a p-dimensional vector which is of form $E[X_{1,1}]$ in which $X_{1,1}$ is a representative of the covariate x (as defined in assumption (B3)) and which is by applying the *binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. Also by the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converges with rate n as $n \to \infty$ to $E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] = \tau$. Thus it holds with (B3) for the second summand of $M^2\phi_2$ that

$$\frac{M^2}{n{\sigma^0}^2} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] - \tau \right) x_{ij} \in \frac{M^2}{n} o(n) = o\left(n^{-\frac{1}{2}} \right).$$

This term converges to zero for $n \to \infty$. By construction of $N_n(\theta^0)$ the random variable $E_{\theta^2} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converges as $n \to \infty$ to $E_{\theta^0} \left[\left(\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} - \tau \right) \varepsilon_{ij} | Y \right] E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$. For this reason and ε_{ij} being iid the third summand of ϕ_3 converges with (B3) as $n \to \infty$ to

$$\frac{1}{n{\sigma^2}^3}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^2}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^2}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}$$
$$-\frac{1}{n{\sigma^0}^3}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}-\tau\right)\varepsilon_{ij}|Y\right]E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}$$
$$\to\left(\frac{1}{{\sigma^2}^3}-\frac{1}{{\sigma^0}^3}\right)E_{\theta^2}\left[E_{\theta^0}\left[\left(\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}-\tau\right)\varepsilon_{1,1}|Y\right]E_{\theta^0}\left[\mathbf{1}_{\{\varepsilon_{1,1}\leqslant 0\}}|Y\right]\right]c_0,$$

which is by applying the *generalised binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant C, where it holds that

$$M^2\frac{M}{K_2}C=\frac{M^3}{K_2}C\leqslant\frac{K_2^{\frac{3}{4}}}{K_2}C=K_2^{-\frac{1}{4}}C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2\phi_3 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.7.4 Case 5: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and can finally conclude

$$M^2\phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.7.5 Case 5: Convergence of ϕ_5

For ϕ_5 see Section 4.2.2.3.5.

4.2.2.8 *Case 6*

By application of Lemma 4.14 the expectation (4.39) in this case can be expressed by

$$\frac{1}{n}E_\theta\left[\frac{\partial^2}{\partial\beta_\tau^2}\ell(\theta|Y)\Big|_{\theta^\star}\right]$$
$$=\frac{1}{n}E_\theta\left[-\frac{1}{\sigma^{\star 2}}\mathbf{X}^T\mathrm{Cov}_{\theta^\star}\left(\mathbf{1}_{\{\varepsilon\leqslant 0\}},\mathbf{1}_{\{\varepsilon\leqslant 0\}}|Y\right)\mathbf{X}\right]$$
$$=-\frac{1}{n\sigma^{\star 2}}\mathbf{X}^T\left(E_\theta\left[E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right.$$
$$\left.-E_\theta\left[E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}|Y\right]E_{\theta^\star}\left[\mathbf{1}_{\{\varepsilon\leqslant 0\}}^T|Y\right]\right]\right)\mathbf{X}$$

$$= -\frac{1}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta} \left[E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij}^T$$
$$+ \frac{1}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta} \left[E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij}^T. \tag{4.52}$$

Whenever $\theta = \theta^\star$ this expression simplifies with $\varepsilon_{ij} \overset{iid}{\sim} ALD(0, \sigma^\star, \tau)$, $E_{\theta^\star} [E_{\theta^\star} [\cdot|Y]] = E_{\theta^\star} [\cdot]$, and Corollary 2.2 to

$$\frac{1}{n} E_{\theta^\star} \left[\frac{\partial^2}{\partial \beta_\tau^2} \ell(\theta|Y) \bigg|_{\theta^\star} \right]$$
$$= -\frac{1}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] x_{ij}^T$$
$$+ \frac{1}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^\star} \left[E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij}^T$$
$$= -\frac{\tau}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} x_{ij}^T$$
$$+ \frac{1}{n\sigma^{\star 2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^\star} \left[E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^\star} \left[\mathbf{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij}^T. \tag{4.53}$$

4.2.2.8.1 Case 6: Convergence of ϕ_1

For ϕ_1 in case 6 I have the expression

$$M^2 \phi_1$$
$$= \frac{M^2}{n} \sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{{\sigma^1}^2} \mathbf{X}^T \mathrm{Cov}_{\theta^1} \left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}, \mathbf{1}_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right.$$
$$\left. + \frac{1}{{\sigma^2}^2} \mathbf{X}^T \mathrm{Cov}_{\theta^2} \left(\mathbf{1}_{\{\varepsilon \leqslant 0\}}, \mathbf{1}_{\{\varepsilon \leqslant 0\}} | Y \right) \mathbf{X} \right)$$
$$= \frac{M^2}{n} \sup_{\theta^1 \in N_n(\theta^0)} \left(-\frac{1}{{\sigma^1}^2} \mathbf{X}^T \left(E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}} \mathbf{1}_{\{\varepsilon \leqslant 0\}}^T | Y \right] \right. \right.$$
$$\left. - E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}} | Y \right] E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T | Y \right] \right) \mathbf{X}$$
$$+ \frac{1}{{\sigma^2}^2} \mathbf{X}^T \left(E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}} \mathbf{1}_{\{\varepsilon \leqslant 0\}}^T | Y \right] \right.$$
$$\left. \left. - E_{\theta^2} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}} | Y \right] E_{\theta^1} \left[\mathbf{1}_{\{\varepsilon \leqslant 0\}}^T | Y \right] \right) \mathbf{X} \right)$$

$$
\begin{aligned}
= \frac{M^2}{n} \sup_{\theta^1 \in N_n(\theta^0)} \Bigg(& -\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \Bigg(\frac{1}{{\sigma^1}^2} x_{ij} E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T \\
& \qquad - \frac{1}{{\sigma^2}^2} x_{ij} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T \Bigg) \\
& + \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \Bigg(\frac{1}{{\sigma^1}^2} x_{ij} E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \\
& \qquad E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T \\
& \qquad - \frac{1}{{\sigma^2}^2} x_{ij} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \\
& \qquad E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T \Bigg) \Bigg),
\end{aligned}
$$

where each summand converges to zero as $n \to \infty$ as shown in the following: By (B4) and the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T$ and $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T$ converge as $n \to \infty$ to $E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] C_1 = \tau C_1$ and $E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} \right] C_1 = \tau C_1$, respectively. C_1 is a $p \times p$-dimensional matrix which is of form $E \left[X_{1,1} X_{1,1}^T \right]$ in which $X_{1,1}$ is a representative of the covariate x (as defined in assumption (B4)). Hence as $n \to \infty$ I have for the first summand of ϕ_1

$$
\begin{aligned}
& \frac{1}{n} \mathbf{X}^T \left(\frac{E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon \leqslant 0\}} \mathbb{1}_{\{\varepsilon \leqslant 0\}}^T \varepsilon | Y \right]}{{\sigma^1}^2} - \frac{E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon \leqslant 0\}} \mathbb{1}_{\{\varepsilon \leqslant 0\}}^T \varepsilon | Y \right]}{{\sigma^2}^2} \right) \mathbf{X} \\
& \quad \to \frac{\tau C_1}{{\sigma^1}^2} - \frac{\tau C_1}{{\sigma^2}^2} = \tau C_1 \left(\frac{1}{{\sigma^1}^2} - \frac{1}{{\sigma^2}^2} \right),
\end{aligned}
$$

which is by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. By the definition of $N_n(\theta^0)$ the random variables $E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ and $E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converge as $n \to \infty$ to $E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$. At the same time by the *Law of Large Numbers* I know that the averaged sum of each independently and identically distributed random variables converges to its expectation and thus I get with (B4)

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} x_{ij} E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^1} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij}^T \\
& \quad \to E_{\theta^0} \left[E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^0} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] C_1
\end{aligned}
$$

and

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] x_{ij}^T$$
$$\to E_{\theta^0}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right] C_1.$$

Hence I have for the second summand of ϕ_1 that

$$\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(\frac{x_{ij} E_{\theta^1}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^1}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] x_{ij}^T}{{\sigma^1}^3} - \frac{x_{ij} E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] x_{ij}^T}{{\sigma^2}^3}\right)$$
$$\to \left(\frac{1}{{\sigma^1}^2} - \frac{1}{{\sigma^2}^2}\right) E_{\theta^0}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right] C_1,$$

which is by (4.49) bounded from above by $\frac{M}{K_2}$ times a constant C. Altogether I have

$$M^2 |\phi_1| \leqslant \frac{M^3}{K_2} C \leqslant \frac{K_2^{\frac{3}{4}}}{K_1} C = K_2^{-\frac{1}{4}} C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2 \phi_1 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.8.2 Case 6: Convergence of ϕ_2

For ϕ_2 in case 6 I get by (4.53) the expression

$$\begin{aligned} M^2\phi_2 = M^2 \Bigg(& -\frac{1}{n{\sigma^2}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] x_{ij}^T \\ & + \frac{1}{n{\sigma^2}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] x_{ij}^T \\ & + \frac{\tau}{n{\sigma^2}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} x_{ij}^T \\ & - \frac{1}{n{\sigma^2}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij} E_{\theta^2}\left[E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right] E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right] x_{ij}^T \Bigg) \end{aligned}$$

$$= M^2 \left(-\frac{1}{n\sigma^{2^2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} x_{ij}^T \right.$$
$$+ \frac{1}{n\sigma^{2^2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right.$$
$$\left. \left. - E_{\theta^2} \left[E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] \right) x_{ij} x_{ij}^T \right).$$

By the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converges with rate n as $n \to \infty$ to $E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{1,1} \leqslant 0\}} \right] = \tau$. Thus it holds with (B4) for the first summand of $M^2 \phi_2$ that

$$\frac{M^2}{n\sigma^{2^2}} \sum_{i=1}^{D} \sum_{j=1}^{n_i} \left(E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] - \tau \right) x_{ij} x_{ij}^T \in \frac{M^2}{n} o(n) \overset{(\star)}{=} o\left(n^{-\frac{1}{2}} \right),$$

where $(\star)$ holds because $M^2 \leqslant K_2^{\frac{1}{2}} = n^{\frac{1}{2}}$ by definition of M – see (4.38). Also by the *Law of Large Numbers* $\frac{1}{n} \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right]$ converges with rate n as $n \to \infty$ to $E_{\theta^2} \left[E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right]$. Thus it holds with (B4) for the second summand of $M^2 \phi_2$ that

$$\frac{M^2}{n\sigma^{2^2}} \left(\sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] x_{ij} x_{ij}^T \right.$$
$$\left. - \sum_{i=1}^{D} \sum_{j=1}^{n_i} E_{\theta^2} \left[E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] E_{\theta^2} \left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}} | Y \right] \right] x_{ij} x_{ij}^T \right)$$
$$\in \frac{M^2}{n} o(n) = o\left(n^{-\frac{1}{2}} \right).$$

Thus this summand also converges to zero as $n \to \infty$, which leads altogether to

$$M^2 \phi_2 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.8.3 Case 6: Convergence of ϕ_3

For ϕ_3 in case 6 I get by (4.52) and (4.53) the expression

$$
\begin{aligned}
&M^2\phi_3 \\
&= M^2\Bigg(-\frac{\tau}{n\sigma^{2\,2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}x_{ij}^T \\
&\quad + \frac{1}{n\sigma^{2\,2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}E_{\theta^2}\left[E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}^T \\
&\quad + \frac{1}{n\sigma^{0\,2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}^T \\
&\quad - \frac{1}{n\sigma^{0\,2}}\sum_{i=1}^{D}\sum_{j=1}^{n_i} x_{ij}E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]x_{ij}^T\Bigg) \\
&= M^2\Bigg(-\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(\frac{\tau}{\sigma^{2\,2}} - \frac{E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]}{\sigma^{0\,2}}\right)x_{ij}x_{ij}^T \\
&\qquad + \frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\Bigg(\frac{E_{\theta^2}\left[E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]}{\sigma^{2\,2}} \\
&\qquad\quad - \frac{E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]}{\sigma^{0\,2}}\Bigg)x_{ij}x_{ij}^T\Bigg).
\end{aligned}
$$

By the *Law of Large Numbers* $\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ converges as $n \to \infty$ to $E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{1,1}\leqslant 0\}}\right] = \tau$. For the first summand of ϕ_3 this leads with (B4) to

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\left(\frac{\tau}{\sigma^{2\,2}} - \frac{E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]\right]}{\sigma^{0\,2}}\right)x_{ij}x_{ij}^T \\
&\quad \to \tau\left(\frac{1}{\sigma^{2\,2}} - \frac{1}{\sigma^{0\,2}}\right)C_1,
\end{aligned}
$$

which is by applying the *binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant. By construction of $N_n(\theta^0)$ the random variable $E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij}\leqslant 0\}}|Y\right]$ converges as

$n \to \infty$ to $E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right] E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right]$. For this reason and ε_{ij} being iid the second summand of ϕ_3 converges with (B4) to

$$\begin{aligned}
&\frac{1}{n{\sigma^2}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right] E_{\theta^2}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right]\right] x_{ij}x_{ij}^T \\
&-\frac{1}{n{\sigma^0}^2}\sum_{i=1}^{D}\sum_{j=1}^{n_i} E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right] E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{ij} \leqslant 0\}}|Y\right]\right] x_{ij}x_{ij}^T \\
&\to \left(\frac{1}{{\sigma^2}^2} - \frac{1}{{\sigma^0}^2}\right) E_{\theta^2}\left[E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{1,1} \leqslant 0\}}|Y\right] E_{\theta^0}\left[\mathbb{1}_{\{\varepsilon_{1,1} \leqslant 0\}}|Y\right]\right] C_1,
\end{aligned}$$

which is by applying the *binomial formula* and (4.49) bounded from above by $\frac{M}{K_2}$ times a constant C, where it holds that

$$M^2\frac{M}{K_2}C = \frac{M^3}{K_2}C \leqslant \frac{K_2^{\frac{3}{4}}}{K_2}C = K_2^{-\frac{1}{4}}C.$$

By assumption on K_2 it holds that $K_2^{-\frac{1}{4}} \to 0$ as $n \to \infty$ and hence

$$M^2\phi_3 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.8.4 Case 6: Convergence of ϕ_4

For ϕ_4 I can use the same argumentations as for ϕ_3 because θ^2 and θ^0 are both in the set $N_n(\theta^0)$ and can finally conclude

$$M^2\phi_4 \to 0 \quad \text{as } n \to \infty.$$

4.2.2.8.5 Case 6: Convergence of ϕ_5

For ϕ_5 see Section 4.2.2.3.5.

4.3 THE CONSISTENCY OF THE PARAMETER ESTIMATOR $\hat{\theta}$

By validating the two assumptions of the *Weiss' Theorem* – see Theorem 4.1 – in Sections 4.2.1 and 4.2.2 I am able to apply the theorem and it follows the existence and the *asymptotic normality* of the parameter estimator $\hat{\theta}(n) = (\hat{\sigma}_V(n), \hat{\sigma}(n), \hat{\beta}_\tau^T(n))^T$ with rate $K(n) = (\sqrt{D}, \sqrt{n}, \sqrt{n}\mathbb{1}_p^T)^T$

$$\begin{pmatrix} \sqrt{D}(\hat{\sigma}_V(n) - \sigma_V^0) \\ \sqrt{n}(\hat{\sigma}(n) - \sigma^0) \\ \sqrt{n}(\hat{\beta}_\tau(n) - \beta_\tau^0) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0, B^{-1}(\theta^0)) \tag{4.54}$$

with the asymptotic covariance matrix $B^{-1}(\theta^0)$. Its inverse is given in Lemma 4.15, where I can see that the entries are not trivial to calculate and as a consequence so is the inverse. Nevertheless the asymptotic covariance matrix is discussed in Section 4.3.2.

4.3.1 *From Asymptotic Normality to Consistency*

The *asymptotic normality* in (4.54) implies the consistency of the parameter estimator vector $\hat{\theta}(n)$: For all $\epsilon > 0$ it holds

$$
\begin{aligned}
& P\left(\left|\hat{\theta}(n) - \theta^0\right| \geqslant \epsilon \mathbf{1}_{p+2}\right) \\
& \quad = P\Big(B^{\frac{1}{2}}(\theta^0)\mathrm{diag}(K(n))\left|\hat{\theta}(n) - \theta^0\right| \\
& \qquad\qquad\qquad\qquad \geqslant B^{\frac{1}{2}}(\theta^0)\mathrm{diag}(K(n))\epsilon\Big) \\
& \quad \overset{(\star)}{\to} 2\left(\mathbf{1}_{p+2} - \phi_{p+2}\left(B^{\frac{1}{2}}(\theta^0)\mathrm{diag}(K(n))\epsilon\right)\right) \\
& \quad \overset{(\star\star)}{\to} 0 \quad \text{as } n \to \infty,
\end{aligned}
$$

where $(\star)$ follows because $B^{\frac{1}{2}}(\theta^0)\mathrm{diag}(K(n))\left|\hat{\theta}(n) - \theta^0\right|$ is by (4.54) *asymptotically* $(p+1)$*-dimensional standard normal distributed* and ϕ_{p+2} is its probability density function. Due to the definition of $K(n)$ tending to ∞ as $n \to \infty$ $(\star\star)$ follows because $\phi_{p+2}(K(n) \cdot C) \to \mathbf{1}_{p+2}$ as $n \to \infty$ for any constant C. As a remark the convergence $(\star)$ is just a between step of the whole convergence proven. This is why at this stage the limit still depends on n.

4.3.2 *The Asymptotic Covariance Matrix*

The asymptotic covariance matrix is given as the inverse of $B(\theta^0)$ in Lemma 4.15. Most of the entries are not analytically calculable but there are a few zeros. Thus all entries but the diagonal in the first column and row are zero. The inverse of a block diagonal matrix can be calculated by taking the inverses of the blocks. Hence the asymptotic covariance matrix $B^{-1}(\theta^0)$ must have zeros at the same entries in first column and row implying the independence of $\hat{\sigma}_V$ to the other estimators. The asymptotic variance of $\sqrt{D}(\hat{\sigma}_V(n) - \sigma_V^0)$ is given by

$$
\left(-\frac{1}{{\sigma_V^0}^2} + \frac{1}{{\sigma_V^0}^6} E\left[E^2\left[V_1^2|Y\right]\right]\Big|_{\theta^0}\right)^{-1} = \left(\frac{-{\sigma_V^0}^4 + E\left[E^2\left[V_1^2|Y\right]\right]\Big|_{\theta^0}}{{\sigma_V^0}^6}\right)^{-1}
$$
$$
= \frac{{\sigma_V^0}^6}{E\left[E^2\left[V_1^2|Y\right]\right]\Big|_{\theta^0} - {\sigma_V^0}^4}.
$$

The other part of the covariance matrix is dependent on expectations of products of conditional expectations. In these conditional expectations I have expressions, which are dependent on the error term $\varepsilon_{1,1}$. Its distribution is the *asymmetric Laplace distribution* $\mathrm{ALD}(0, \sigma, \tau)$, which has in its density the scale $\tau(1-\tau)$. In linear quantile models this scale is part of the asymptotic covariance matrix as discussed in Section 2.5. Thus it will be natural that its inverse $\frac{1}{\tau(1-\tau)}$ is part in $B^{-1}(\theta^0)$ in the variance part of $\hat{\beta}_\tau$. As a result the estimation for $\tau = 0.5$ must have the smallest asymptotic variance for $\hat{\beta}_\tau$. This will be discussed in the following Section 4.4. The part C_1 can be interpreted of form $E\left[X_{1,1}X_{1,1}^T\right]$ which is the second moment of the independent variable $X_{1,1}$ and therefore a representative of the variance in $X_{1,1}$. Hence for an intercept estimation it is 1 and has no impact on the asymptotic variance. For all others the variance in the independent variable impacts the asymptotic variance of $\hat{\beta}_\tau$.

4.4 SIMULATION STUDY OF THE ASYMPTOTIC NORMALITY

In this simulation study I am going to investigate the asymptotic behaviour of the distribution of the estimators $\hat{\sigma}_V$, $\hat{\sigma}$, and $\hat{\beta}_\tau$ as well as the asymptotic variances and covariances of and between them, respectively. Thus we are able to see the proven asymptotic normality and the convergence rates. Furthermore we will get an idea of the dependence of the asymptotic variances and covariances on τ.

4.4.1 *The Setup*

In a simulation study I produced each 500 pseudo populations with 500 areas with $N_i = 200$ individuals each ($i = 1, 2, \ldots, 500$). The model used for $\tau = 0.6$ and $\tau = 0.9$ is

$$Y_{ij} = 2 + 0.8x_{ij} + V_i + \varepsilon_{ij},$$

where the independent variables x_{ij} come from a *uniform distribution* on $(0, 1)$

$$x_{ij} \sim U(0, 1), \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots, 200.$$

Note that in this setting β_τ is the same for both τ. Thus $\beta_{0.6} = \beta_{0.9} = (2, 0.8)^T$. The random effect was drawn from a *normal distribution* with zero mean and variance $\sigma_V^2 = 0.3^2$ for each τ

$$V_i \sim N(0, 0.3^2), \quad i = 1, 2, \ldots, 500,$$

and the error term was drawn from an *asymmetric Laplace distribution* with scale parameter $\sigma = 0.5$ and $\tau = 0,6$ or $\tau = 0.9$

$$\varepsilon_{ij} \sim \mathrm{ALD}(0, 0.5, \tau), \quad i = 1, 2, \ldots, 500; j = 1, 2, \ldots, 200.$$

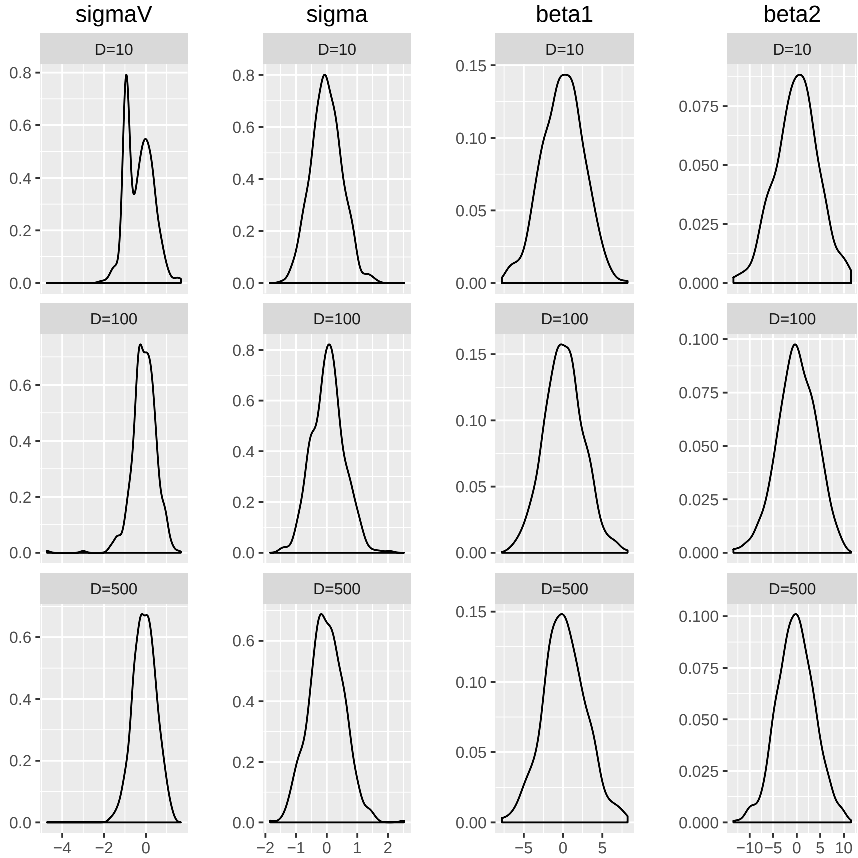

Figure 4.1: Densities of $K_\iota(n)(\hat{\theta}_\iota(n) - \theta^0)$ in 500 estimations of $\hat{\sigma}_V$, $\hat{\sigma}$, $\hat{\beta}_1$, and $\hat{\beta}_{\tau,2}$, $D = 10, 100, 500$, and $\tau = 0.6$

4.4.2 *The Convergence to Normality*

In each area $n_i = 10$ observations were drawn and the number of areas D in the three cases varied from 10 over 100 to 500. The estimation was fulfilled with the software R and the package lqmm by Geraci [2016]. The details of the methods used in this package are topics of Geraci and Bottai [2014]. In Figures 4.1 and 4.2 the densities of the 500 transformed estimators in the manner

$$K_\iota(n)(\hat{\theta}_\iota(n) - \theta^0)$$

are drawn for the four estimators $\hat{\sigma}_V$, $\hat{\sigma}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ in the three cases of $D = 10, 100, 500$. As a reminder the convergence rates for the estimators were $K_1(n) = \sqrt{D}$ and $K_\iota(n) = \sqrt{n} = \sqrt{10}\sqrt{D}$ for $\iota = 2, 3, 4$ such that they are $K_1(n) = \sqrt{10}, 10, \sqrt{5} \cdot 10$ and $K_\iota(n) = \sqrt{n} = 10, \sqrt{10} \cdot 10, \sqrt{50} \cdot 10$ for $\iota = 2, 3, 4$ for the three cases. One can see that in both different cases for τ the densities of the distribution of all parameter estimators converge to the shape of a *normal distribution*. This happens faster for the three last parameters $\hat{\sigma}$, $\hat{\beta}_1$, and $\hat{\beta}_{\tau,2}$

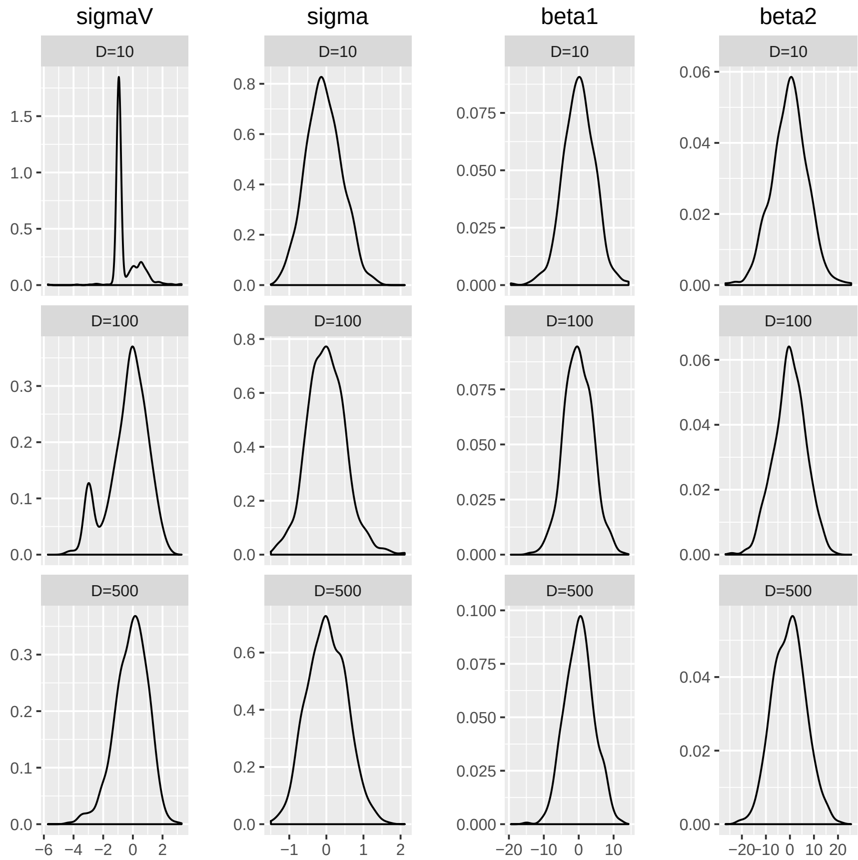

Figure 4.2: Densities of $K_\iota(n)(\hat{\theta}_\iota(n) - \theta^0)$ in 500 estimations of $\hat{\sigma}_V$, $\hat{\sigma}$, $\hat{\beta}_1$, and $\hat{\beta}_{\tau,2}$, $D = 10, 100, 500$, and $\tau = 0.9$

due to the faster convergence (by $\sqrt{n_i} = \sqrt{10}$ here). However for all parameter estimators the distribution has the same support over all different area numbers D. This validates the convergence rates afterwards. Another observation is that the support for $\hat{\sigma}_V$ and $\hat{\sigma}$ is about the same for $\tau = 0.6$ and $\tau = 0.9$. This implies that τ has no impact on the asymptotic variance, which can not be quantified due to the unknown measure $P_{V|Y}$. In contrast the support for $\hat{\beta}_{\tau,1}$ and $\hat{\beta}_{\tau,2}$ almost doubles for $\tau = 0.9$. Thus τ must have an impact on the asymptotic variance. This is an intuitive observation because quantile estimators for τ distant to 0.5 must have a worse power due to the leverage of observations Y_{ij} on the edges. Nevertheless it is positive that τ seems to have no impact on the variance estimation in $\hat{\sigma}_V$ and $\hat{\sigma}$.

4.4.3 *The Asymptotic Variances and Covariances*

Since the asymptotic covariance matrix cannot be given analytically, I observe the empirical variances and covariances now. In Table 4.1 all empirical variances and covariances between the estimators σ_V, σ,

$\tau = 0.6$		$\hat{\sigma}_V$	$\hat{\sigma}$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
	D=10	0.13183383			
$\hat{\sigma}_V$	D=100	0.05115810			
	D=500	0.02706904			
	D=10	-0.0122787603	0.013020581		
$\hat{\sigma}$	D=100	-0.0028792905	0.001507195		
	D=500	-0.0009334091	0.000310581		
	D=10	-0.014148793	0.000234930	0.41593651	
$\hat{\beta}_{\tau,1}$	D=100	0.001343279	0.000089589	0.02817245	
	D=500	-0.001034811	0.000059185	0.00562185	
	D=10	0.001384736	-0.0018889107	-0.630198698	1.33194847
$\hat{\beta}_{\tau,2}$	D=100	-0.006078426	0.0000636330	-0.043792012	0.09063374
	D=500	0.001243569	-0.0001072401	-0.008387372	0.01700046
$\tau = 0.9$		$\hat{\sigma}_V$	$\hat{\sigma}$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
	D=10	0.08059054			
$\hat{\sigma}_V$	D=100	0.07390737			
	D=500	0.02888356			
	D=10	-0.004020231	0.0115743598		
$\hat{\sigma}$	D=100	-0.002250779	0.0011812333		
	D=500	-0.000889200	0.0002748655		
	D=10	-0.023446911	-0.0047595936	1.56164470	
$\hat{\beta}_{\tau,1}$	D=100	-0.017782719	0.0010442254	0.10024339	
	D=500	-0.004507479	0.0002003062	0.01653899	
	D=10	-0.0195119402	0.019300360	-2.56110461	5.27421811
$\hat{\beta}_{\tau,2}$	D=100	0.0046714403	-0.000157490	-0.14932187	0.30439060
	D=500	-0.0000417081	0.000092882	-0.02379976	0.04907912

Table 4.1: Empirical variances and covariances between estimators in 500 estimations of $\hat{\sigma}_V$, $\hat{\sigma}$, $\hat{\beta}_{\tau,1}$, and $\hat{\beta}_{\tau,2}$, $D = 10, 100, 500$ for $\tau = 0.6$ and $\tau = 0.9$

$\beta_{\tau,1}$, and $\beta_{\tau,2}$ are displayed for $\tau = 0.6$ and $\tau = 0.9$ and the different sample sizes $D = 10, 100, 500$ or $n = 100, 1000, 5000$.

4.4.3.1 *The Asymptotic Variances*

We can see that all variances and covariances decrease with larger D or n. As already seen in the densities the convergence rates can be observed. Thus the variance for $\hat{\sigma}_V$ in $\tau = 0.6$ is decreasing from

$\tau = 0.6$		$\hat{\sigma}_V$	$\hat{\sigma}$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
	D=10	1.318338			
$\hat{\sigma}_V$	D=100	5.115810			
	D=500	13.534518			
	D=10	-0.3882885	1.302058		
$\hat{\sigma}$	D=100	-0.9105116	1.507195		
	D=500	-1.4758494	1.552905		
	D=10	-0.4474241	0.23493043	41.59365	
$\hat{\beta}_{\tau,1}$	D=100	0.4247820	0.08958872	28.17245	
	D=500	-1.6361799	0.29592235	28.10925	
	D=10	0.04378921	-0.18889069	-63.01987	133.19485
$\hat{\beta}_{\tau,2}$	D=100	-1.92216714	0.06363633	-43.79201	90.63374
	D=500	1.96625548	-0.53620052	-41.93686	85.00230
$\tau = 0.9$		$\hat{\sigma}_V$	$\hat{\sigma}$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
	D=10	0.8059054			
$\hat{\sigma}_V$	D=100	7.3907368			
	D=500	14.4417822			
	D=10	-0.3882885	1.157436		
$\hat{\sigma}$	D=100	-0.9105116	1.181233		
	D=500	-1.4758494	1.374328		
	D=10	-0.7414564	-0.4759594	156.16447	
$\hat{\beta}_{\tau,1}$	D=100	-5.6233894	1.0442254	100.24339	
	D=500	-7.1269499	1.0015311	82.69494	
	D=10	-0.61702173	1.9300356	-256.1105	527.4218
$\hat{\beta}_{\tau,2}$	D=100	1.47723913	-0.1574949	-149.3219	304.3906
	D=500	-0.06594629	0.000092882	-118.9988	245.3956

Table 4.2: By convergence rate corrected empirical variances and covariances between estimators in 500 estimations of $\hat{\sigma}_V$, $\hat{\sigma}$, $\hat{\beta}_{\tau,1}$, and $\hat{\beta}_{\tau,2}$, $D = 10, 100, 500$ for $\tau = 0.6$ and $\tau = 0.9$

0.13183383 over 0.05115810 to 0.02706904, where it holds for the by the convergence rate corrected variances in Table 4.2

$$10 \cdot 0.13183383 = 1.318338$$
$$100 \cdot 0.05115810 = 5.115810$$
$$500 \cdot 0.02706904 = 13.534518.$$

In theory these numbers must be approximately equal. One reason why this is not observable here may be that D is too small. As seen in the densities in Figures 4.1 and 4.2 the convergence to *normality* is slow and seems not to be reached yet. For $\tau = 0.9$ I get a similar

picture with by the convergence rate corrected variances of 0.8059054, 7.3907368, and 14.4417822 (see Table 4.2). Nevertheless the numbers are about the same for the same D but different τ, which implies that τ has no impact on the estimation of σ_V.
Different to that is the behaviour of the by the convergence rate corrected variances of the three other parameter estimators $\hat{\sigma}$, $\hat{\beta}_{\tau,1}$, and $\hat{\beta}_{\tau,2}$. For $\hat{\sigma}$ in the case $\tau = 0.6$ I have in Table 4.2 corrected variances of

$$\begin{aligned} 100 \cdot 0.013020581 &= 1.302058 \\ 1000 \cdot 0.001507195 &= 1.507195 \\ 5000 \cdot 0.000310581 &= 1.552905. \end{aligned}$$

For $\tau = 0.9$ I have 1.157436, 1.181233, and 1.374328 (see Table 4.2). These results are approximately equal over all D and τ, which proves the convergence rate of $\sqrt{n}$ and the independence of τ in the estimation of σ. For $\hat{\beta}_1$ in the case $\tau = 0.6$ I have for $n = 100, 1000, 5000$ the corrected covariances 41.59365, 28.17245, and 28.10925 and for $\tau = 0.9$ I have 156.16447, 100.24339, and 82.69494 (see Table 4.2). For each τ the results are about the same, which verifies the convergence rate of $\sqrt{n}$. Nevertheless the values for different τ differ. Thus the variance increases for τ distant to 0.5. As mentioned in Section 4.3.2 the variances may have the scale of $\frac{1}{\tau(1-\tau)}$, which can be seen here. By correcting the variances of $\hat{\beta}_{\tau,1}$ with respect to the inverse scale $\tau \cdot (1-\tau)$ I have

$$\begin{aligned} D = 10: &\quad 0.6 \cdot 0.4 \cdot 0.41593651 = 0.099824763 \\ &\quad 0.9 \cdot 0.1 \cdot 1.56164470 = 0.140548023 \\ D = 100: &\quad 0.6 \cdot 0.4 \cdot 0.02817245 = 0.006761389 \\ &\quad 0.9 \cdot 0.1 \cdot 0.10024339 = 0.009021905 \\ D = 500: &\quad 0.6 \cdot 0.4 \cdot 0.00562185 = 0.001349244 \\ &\quad 0.9 \cdot 0.1 \cdot 0.01653899 = 0.001488509. \end{aligned}$$

I can see the convergence to the scale of $\frac{1}{\tau(1-\tau)}$ in the estimation of $\beta_{\tau,1}$ because the values for the same D are about equal for both τ.
For $\hat{\beta}_{\tau,2}$ in the cases $\tau = 0.6$ and $\tau = 0.9$ the corrected variances I have in Table 4.2 are approximately equal over all $n = 100, 1000, 5000$, which proves the convergence rate of $\sqrt{n}$. By multiplying the variances with the inverse scale of $\tau(1-\tau)$ I get

$$\begin{aligned} D = 10: &\quad 0.6 \cdot 0.4 \cdot 1.33194847 = 0.31966763 \\ &\quad 0.9 \cdot 0.1 \cdot 5.27421811 = 0.474679630 \\ D = 100: &\quad 0.6 \cdot 0.4 \cdot 0.09063374 = 0.02175210 \\ &\quad 0.9 \cdot 0.1 \cdot 0.30439060 = 0.027395154 \end{aligned}$$

$$D = 500: \qquad 0.6 \cdot 0.4 \cdot 0.01700046 = 0.00408011$$
$$0.9 \cdot 0.1 \cdot 0.04907912 = 0.004417121.$$

The convergence to the scale of $\frac{1}{\tau(1-\tau)}$ in the estimation of $\beta_{\tau,2}$ is also visible here because the values for the same D are about equal for both τ with increasing D. In all cases of D and τ the variance is bigger than the one of $\hat{\beta}_{\tau,1}$. This can be explained with the variance of X, which is added here.

4.4.3.2 *The Asymptotic Covariances*

Let me now observe the covariances. As an analytical result the covariances between $\hat{\sigma}_V$ and all other estimators must converge to zero. By correcting the results in Table 4.1 with respect to the convergence rates I get for the corrected covariances between $\hat{\sigma}_V$ and $\hat{\sigma}$ (see Table 4.2)

$$\tau = 0.6: \qquad \sqrt{10} \cdot \sqrt{100} \cdot -0.0122787603 = -0.3882885$$
$$\sqrt{100} \cdot \sqrt{1000} \cdot -0.0028792905 = -0.9105116$$
$$\sqrt{500} \cdot \sqrt{5000} \cdot -0.0009334091 = -1.4758494$$
$$\tau = 0.9: \qquad \sqrt{10} \cdot \sqrt{100} \cdot -0.004020231 = -0.1271309$$
$$\sqrt{100} \cdot \sqrt{1000} \cdot -0.002250779 = -0.7117587$$
$$\sqrt{500} \cdot \sqrt{5000} \cdot -0.000889200 = -1.4059487.$$

These results do not only contradict the analytics but also imply a negative covariance between the variance parameter estimators. This may be explained by the not yet convergence of the estimator $\hat{\sigma}_V$ on the one hand. The correlations for $\tau = 0.6$ are $-0.2963643, -0.3279016, -0.3219202$ and for $\tau = 0.9$ they are $-0.1316317, -0.2408913, -0.3155833$. The correlation corrects the covariance by the standard deviations and the convergence rates are cancelled. As a result the correlations are of same order for every D and τ. Nevertheless it is negative, which may be explained by the estimation process on the other hand. Since the roots of the first derivatives cannot be calculated analytically, there is a Gaussian quadrature formula used in the lqmm package. The negative sign implies that $\hat{\sigma}_V$ and $\hat{\sigma}$ are negatively correlated. The two variance estimators divide the overall variance. Whenever $\hat{\sigma}_V$ increases $\hat{\sigma}$ decreases and vice versa.

The correlations between $\hat{\sigma}_V$ and $\hat{\beta}_{\tau,1}$ for $D = 10, 100, 500$ are

$$\tau = 0.6: \qquad -0.06042163, 0.03538315, -0.08388514$$
$$\tau = 0.9: \qquad -0.06609255, -0.20659815, -0.20623095.$$

$\tau = 0.6$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
D=10	0.03192350	-0.014343385
D=100	0.01374853	0.005444718
D=500	0.04478996	-0.046670177
$\tau = 0.9$	$\hat{\beta}_{\tau,1}$	$\hat{\beta}_{\tau,2}$
D=10	-0.03540220	0.07811560
D=100	0.09596184	-0.00830583
D=500	0.09394634	0.02528847

Table 4.3: Correlations between estimators in 500 estimations of $\hat{\sigma}$ with $\hat{\beta}_{\tau,1}$ and $\hat{\beta}_{\tau,2}$, $D = 10, 100, 500$ for $\tau = 0.6$ and $\tau = 0.9$

The correlations between $\hat{\sigma}_V$ and $\hat{\beta}_{\tau,2}$ for $D = 10, 100, 500$ are

$$\begin{aligned} \tau = 0.6: &\quad 0.003304532, -0.089266599, 0.057969989 \\ \tau = 0.9: &\quad -0.029928117, 0.031145214, -0.001107762. \end{aligned}$$

These correlations are close to zero and therefore verify the results in Lemma 4.15 and the further thoughts in Section 4.3.2.
The corrected covariances between $\hat{\sigma}$ and $\hat{\beta}_{\tau,1}$ in Table 4.2 with respect to the convergence rates are

$$\begin{aligned} \tau = 0.6: &\quad 100 \cdot 0.002349304 = 0.23493043 \\ &\quad 1000 \cdot 0.00008958872 = 0.08958872 \\ &\quad 5000 \cdot 0.00005918447 = 0.29592235 \\ \tau = 0.9: &\quad 100 \cdot -0.0047595936 = -0.4759594 \\ &\quad 1000 \cdot 0.0010442254 = 1.0442254 \\ &\quad 5000 \cdot 0.0002003062 = 1.0015311. \end{aligned}$$

For the corrected covariances between $\hat{\sigma}$ and $\hat{\beta}_{\tau,2}$ I have for $\tau = 0.6$ and $\tau = 0.9$ similar values. The correlations between $\hat{\sigma}$ and $\hat{\beta}_{\tau,1}$ or $\hat{\beta}_{\tau,2}$ are given in Table 4.3. They are all close to zero, which implies no correlations between the parameter estimates. This was not analytically seen before and therefore it is an interesting result.
The corrected covariances between $\hat{\beta}_{\tau,1}$ and $\hat{\beta}_{\tau,2}$ in Table 4.2 already have the same sign and are of same order. For the empirical correlations I get for $n = 100, 1000, 5000$

$$\begin{aligned} \tau = 0.6: &\quad -0.8466819, -0.8666378, -0.8579380 \\ \tau = 0.9: &\quad -0.8923951, -0.8548304, -0.8353524. \end{aligned}$$

Hence they are highly negatively correlated with an average correlation around -0.85 for all D and τ. This was not observable before since $B^{-1}(\theta^0)$ is not analytically calculable.

4.4.4 *The Conclusion of the Simulation Study*

As a result of the simulation study I can state that the *asymptotic normality* with the given rates can be observed. The convergence for the estimators of σ, $\beta_{\tau,1}$, and $\beta_{\tau,2}$ is faster than the one for σ_V, which is justified by the faster rate of $\sqrt{n}$ compared to $\sqrt{D}$.
Furthermore I was able to discuss the covariance structure. The variances of the variance estimators $\hat{\sigma}_V$ and $\hat{\sigma}$ are quite small and do not depend on τ. The variance of the estimator $\hat{\beta}_\tau$ is quite big and depends on τ with rate $\frac{1}{\tau(1-\tau)}$. Thus for τ close to 0.5 the variances are minimal. As a surprise the correlation between $\hat{\sigma}_V$ and $\hat{\sigma}$ seems to be negative, although it must converge to zero in theory. This can be due to the numerical estimation process. On the other hand the observed correlation between $\hat{\sigma}$ and $\hat{\beta}_{\tau,1}$ or $\hat{\beta}_{\tau,2}$ is around zero for all τ, which implies the asymptotic independence between those estimators. Similar as in any regression the parameter estimators $\beta_{\tau,1}$ and $\beta_{\tau,2}$ are highly correlated.

4.5 CONCLUSION

This Chapter contains the major part the main result of this thesis. I was able to show the *asymptotic normality* of the parameters estimated in the linear quantile mixed model (4.25) under the stated assumptions (B1) to (B6).
This undertaking was oriented on the approach by Miller [1977] and Pinheiro [1994], which were an article and a thesis proofing the *asymptotic normality* of the parameter estimators in the linear mixed model (3.1) with *normal* error terms. Similar to their approach I was able to apply the *Weiss' Theorem* – see Theorem 4.1 – for non-standard cases. In order to show the assumptions of this theorem I had to develop the second derivatives of the log-likelihood density $\ell(\theta|Y)$ and prove two main assumptions on them. Since the density of the observations Y was only given as a marginal density in integral form, the calculation of the second derivatives was quite complicated. After all I was able to express the results in forms of conditional expectations. This field of conditional expectation helped with the further handling. Indeed the further calculations did not become less complicated but I was able to apply results for conditional covariances and variances.
In the end of Section 4.2.1 the inverse of the asymptotic covariance matrix of the parameter estimators was stated in Lemma 4.15. I was able to show the continuity and positive definiteness of this matrix, which proved Assumption 1 of the *Weiss' Theorem*. The asymptotic covariance matrix was unfortunately not analytically calculable. However I was able to discuss it in a small simulation study in Section 4.4. The second assumption of the theorem was more or less a matter of construction. In Section 4.2.2 I mainly oriented on the outline in

Pinheiro [1994]. In the quantile approach it was however less clear at some points.

After all the *Weiss' Theorem* was applicable and the *asymptotic normality* and the convergence rates were the results for the parameter estimation. In Section 4.3 I showed that the *asymptotic normality* implies the consistency of the estimators. This on the other hand is a main part in the proof of the consistency of the quantile estimation in linear mixed models as shown in Section 3.4.

5 SUMMARY

5.1 CONCLUSION

In summary of this work I can state, that I was able to show the consistency of quantile estimation in linear mixed models under the mentioned assumptions. This undertaking turned out to be rather extensive due to the complex appearance of the log-likelihood density (cf. Chapters 3 and 4). Nevertheless I was able to apply the *Weiss' Theorem* for the *asymptotic normality* of the *maximum likelihood* parameter estimator in the first step of the quantile estimation. The same theorem was already employed in the proof of the consistency of the mean estimation in linear mixed models (cf. Miller [1977] and Pinheiro [1994]). In the linear quantile mixed model I firstly utilised this statement in order to show the first part of the consistency proof of the estimator. Since this proof is extensive, I outsourced it into the last chapter – Chapter 4 – of this thesis. By doing so I want to emphasise this main achievement. A model-based simulation study completes and verifies the analytical approach of this Chapter 4.
In comparison to the first part of the consistency proof in the linear quantile model, the second part was relatively compact. Therefore it is directly shown in the main theorem of this dissertation – Theorem 3.1. There I needed some statements about matrices of algebraic manner, which can be found in the Appendix A. However the consistency of the quantile estimator in linear mixed models is proven and is further supported by simulation studies in Section 3.4.2. This result contributes an asymptotic theory to the quantile regression in the field of linear mixed models. It also includes a discussion about the *mean squared error* (MSE) of the estimator in Section 3.5.
Furthermore I showed that the linear quantile mixed model can also be applied to count data (cf. Section 3.6.1). This approach is based on the count quantile method in linear quantile models by Machado and Santos Silva [2005]. Moreover I showed that the consistency also holds in the count case.
In Section 3.6.2 I also introduced a method, called *Microsimulation via Quantiles*, predicated on the quantile estimation in linear mixed models. This methodology gives the ability to estimate any parameter of interest of a population. It is founded on the natural interconnection between quantiles and the distribution function. From the estimation of quantiles for a given grid of τ an empirical distribution function can be derived. Using a *Monte Carlo* simulation or microsimulation of the population the estimate of the parameter of interest can be

calculated. *Microsimulation via Quantiles* is a tool for area or group estimates, which are beyond the mean. In mean estimation one can take advantage of the linearity of means. E.g. when we are interested in an area mean, we take the mean of all unit means. For other parameter, for example area quantiles, this linearity does not hold. In a simulation study I showed first results of this approach for the estimation of area quantiles. Still the method can be improved by further corrections on e.g. quantile crossing or the bias. Meanwhile the asymptotic discussion in this thesis could be used as a starting point for the bias correction.

5.2 OUTLOOK

Further research can be continued at two results of this dissertation. The first is the main result, the consistency of the quantile estimator in linear mixed models, and the second one is the idea of estimating area parameters beyond the mean using *Microsimulation via Quantiles*. The consistency of the quantile estimator is an important and wishful property. It implies the asymptotic unbiasedness and the convergence of the *mean squared error* to zero. However there are further research questions on the *mean squared error* of the estimator. When, in practice, the sample size is finite, one may be interested in analytical and numerical approaches on the estimation of the *mean squared error*. This may also lead in a bias-corrected version of the proposed quantile estimator.

Furthermore the quantile estimation in linear mixed models can be discussed in focus of problems, which already appeared in the quantile regression in linear models. The application to count data is just one extension. One may also consider the application in the context of time series data. In addition the problem of quantile crossing can also be observed in linear mixed models. A solution for this question is outstanding. On these extensions an asymptotical theory can be built. The methodology in this thesis can be used as a starting point addressing these issues.

The idea of *Microsimulation via Quantiles* can also be extended. Further discussion and simulations can be found in the working paper by Weidenhammer et al. [2016]. It is just intuitive to bring the idea of count quantiles and *Microsimulation via Quantiles* together. Then also area parameters of interest can be estimated for discrete data. The method allows the estimation of area medians of continuous and discrete distributions and presents a robust alternative to area mean estimators. Further parameters like poverty indicators or proportions can be examined.

In context of the *Microsimulation via Quantiles* one can also discuss the effect of quantile crossing on the estimate and the method can be adapted accordingly in these cases. Also a bias-correction in the

quantile estimation may improve the method.
Last but not least the method *Microsimulation via Quantiles* itself needs an asymptotic theory. In this regard this dissertation may also be a starting point. As a reminder the quantile estimation here was proven for a single and fixed τ. Thus it requires an extension to a grid of τ. Once a set of quantile estimators leads to a consistently estimated distribution function, the consistency of the parameter estimate can be derived with *Monte Carlo* approaches for asymptotic behaviour.

A

APPENDIX

The following three lemmata have a linear algebra character. Since the form of the following matrix $\mathbf{X}$ is not common and has rather a special form the proofs of the results are exercises I would like to execute.

Lemma A.1. *For a symmetric matrix* $\mathbf{X} \in \mathbb{R}^{n \times n}$ *of form*

$$\mathbf{X} = a\mathrm{I}_n + b\mathbf{1}_{n\times n} = \begin{pmatrix} a+b & b & b & \dots & b \\ b & a+b & b & \dots & b \\ & & & & \\ b & b & \dots & b & a+b \end{pmatrix}$$

with $a \in \mathbb{R} \setminus \{0\}$ *and* $b \in \mathbb{R} \setminus \left\{-\frac{a}{n}\right\}$ *the inverse matrix* $\mathbf{X}^{-1}$ *exists and it holds*

$$\mathbf{X}^{-1} = c\mathrm{I}_n + d\mathbf{1}_{n\times n} = \begin{pmatrix} c+d & d & d & \dots & d \\ d & c+d & d & \dots & d \\ & & & & \\ d & d & \dots & d & c+d \end{pmatrix}$$

with

$$c = \frac{1}{a} \quad \textit{and}$$
$$d = -\frac{b}{a(a+nb)}.$$

Proof. The matrix product of $\mathbf{X}$ and $\mathbf{X}^{-1}$ given in the Lemma can be calculated as

$$\begin{aligned} \mathbf{X}\mathbf{X}^{-1} &= (a\mathrm{I}_n + b\mathbf{1}_{n\times n})(c\mathrm{I}_n + d\mathbf{1}_{n\times n}) \\ &= ac\mathrm{I}_n + (bc + ad)\,\mathbf{1}_{n\times n} + bd\mathbf{1}_{n\times n}\mathbf{1}_{n\times n}. \end{aligned}$$

For the matrix product of two $n \times n$-dimensional matrices of ones it holds

$$\mathbf{1}_{n\times n}\mathbf{1}_{n\times n} = n\mathbf{1}_{n\times n}$$

and thus

$$\mathbf{X}\mathbf{X}^{-1} = ac\mathrm{I}_n + (bc + ad + nbd)\,\mathbf{1}_{n\times n}.$$

By setting c and d as stated in the Lemma I get

$$\begin{aligned}\mathbf{X}\mathbf{X}^{-1} &= a\frac{1}{a}I_n + \left(b\frac{1}{a} + a\left(-\frac{b}{a(a+nb)}\right) + nb\left(-\frac{b}{a(a+nb)}\right)\right)\mathbf{1}_{n\times n} \\ &= I_n + \frac{b(a+nb) - ab - nb^2}{a(a+nb)}\mathbf{1}_{n\times n} \\ &= I_n.\end{aligned}$$

Hence $\mathbf{X}^{-1}$ is the inverse matrix of $\mathbf{X}$. Its existence is given for all $a \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R} \setminus \left\{-\frac{a}{n}\right\}$. □

Lemma A.2. *The sum of all entries of the inverse matrix* $\mathbf{X}^{-1}$ *is given as*

$$\sum_{i,j=1}^{n} x_{ij}^{-1} = \frac{n}{a+nb}.$$

For $b \neq 0$ *and* $n \to \infty$ *the sum converges to* $\frac{1}{b}$

$$\lim_{n\to\infty} \sum_{i,j=1}^{n} x_{ij}^{-1} = \frac{1}{b}.$$

Proof. The sum of all entries of $\mathbf{X}^{-1}$ can be calculated as follows

$$\sum_{i,j=1}^{n} x_{ij}^{-1} = nc + n^2 d.$$

With $c = \frac{1}{a}$ and $d = -\frac{b}{a(a+nb)}$ it develops to

$$\begin{aligned}\sum_{i,j=1}^{n} x_{ij}^{-1} &= n\frac{1}{a} + n^2\left(-\frac{b}{a(a+nb)}\right) \\ &= \frac{n(a+nb) - n^2 b}{a(a+nb)} \\ &= \frac{na}{a(a+nb)} \\ &= \frac{n}{a+nb}.\end{aligned}$$

For $b \neq 0$ and $n \to \infty$ I can apply the rule of L'hôpital $(\star)$ and get for the limit

$$\begin{aligned}\lim_{n\to\infty} \sum_{i,j=1}^{n} x_{ij}^{-1} &= \lim_{n\to\infty} \frac{n}{a+nb} \\ &\overset{(\star)}{=} \lim_{n\to\infty} \frac{1}{b} \\ &= \frac{1}{b}.\end{aligned}$$

□

Lemma A.3. *The column sum for each column* $j = 1, 2, \ldots, n$ *of the inverse matrix* $\mathbf{X}^{-1}$ *is given as*

$$\sum_{i=1}^{n} x_{ij}^{-1} = \frac{1}{a + nb}.$$

For $n \to \infty$ *the column sums converge to zero*

$$\lim_{n \to \infty} \sum_{i=1}^{n} x_{ij}^{-1} = 0.$$

Proof. All column sums of $\mathbf{X}^{-1}$ can be calculated as follows

$$\sum_{i=1}^{n} x_{ij}^{-1} = c + nd.$$

With $c = \frac{1}{a}$ and $d = -\frac{b}{a(a+nb)}$ it develops to

$$\begin{aligned} \sum_{i=1}^{n} x_{ij}^{-1} &= \frac{1}{a} + n\left(-\frac{b}{a(a+nb)}\right) \\ &= \frac{(a+nb) - nb}{a(a+nb)} \\ &= \frac{1}{a+nb}. \end{aligned}$$

For $n \to \infty$ I get for the limit

$$\lim_{n \to \infty} \sum_{i=1}^{n} x_{ij}^{-1} = \lim_{n \to \infty} \frac{1}{a+nb} = 0.$$

□

ABSTRACT

Quantiles are parameters of a distribution, which are of location and of scale character at the same time. The median, as a location parameter, is even robust and outperforms the mean, whenever there are outliers or extreme values in the data. In linear models quantile regression was firstly introduced by Koenker and Bassett [1978]. The method is well investigated and the asymptotic behaviour, like the consistency, was already proven. Doing so one is able to use the equivalence of the linear quantile model to a linear model with asymmetric Laplacian error terms. The asymmetric Laplace distribution has established a direct link to quantile estimation and is investigated in Yu and Zhang [2005].

In linear mixed models the quantile estimation was recently developed by Geraci and Bottai [2007] as well as Geraci and Bottai [2014]. There also the equivalence to an asymmetric Laplacian mixed model is employed. The estimation of the quantile is possible due to the shift to a maximum likelihood approach. An estimating algorithm of numerical kind is implemented in the open software R (see the package lqmm by Geraci [2016]). Due to the complex appearance of the log-likelihood density analytical solutions are pending. However the asymptotic theory was outstanding, up to this thesis, which shows the consistency of the conditional quantile estimator under some additional conditions. In proofing this property the Weiss' Theorem (cf. Weiss [1971] and Weiss [1973]) for the dependent observations in the maximum likelihood estimation is applied. In the linear mixed model Miller [1977] and Pinheiro [1994] also employed this theorem, when they proved the asymptotic normality of the parameter estimators for the mean estimation. In the quantile estimation the necessity for its application is the calculation of the second derivatives from the log-likelihood density with respect to the unknown parameters. These constitute a form of the Fisher information matrix. Therefore they determine the asymptotic variance of the parameter estimators and are needed for the proof of the assumption in the Weiss' Theorem. The resulting asymptotic normality of the parameter estimators imply the consistency of a conditional τ-quantile estimator for a given value $\tau \in (0, 1)$. Both proven properties, the asymptotic normality of the parameter estimators and the consistency of the quantile estimator, are supported by model-based simulation studies.

The application of quantile regression in linear mixed models is shown to be applicable for count data. In this special case the consistency is also proven here. Furthermore a method called Microsimulation via Quantiles for the estimation of parameters, which are beyond

the mean of a population, is proposed. There the natural connection between quantiles and the distribution function is deployed leading into an estimation of the whole distribution. From there any parameter of interest – e.g. be a quantile, a proportion, or others – can be generated by a Monte Carlo simulation or microsimulation.

ZUSAMMENFASSUNG

Quantile sind Kennzahlen einer Verteilung, welche sowohl die Lage als auch die Streuung dieser darstellen. Der Median, als Lageparameter, ist dabei robust und übertrifft das arithmetische Mittel, wenn Ausreißer oder Extremwerte in den Beobachtungen vorliegen. Im linearen Modell wurde die Quantilsregression zuerst von Koenker and Bassett [1978] entwickelt. Die Methodik ist gut erforscht und das asymptotische Verhalten, wie die Konsistenz, wurde bereits bewiesen. Dabei kann die Äquivalenz des linearen quantilen Modells zu einem linearen Modell mit asymptotisch Laplace-verteilten Fehlertermen genutzt werden. Die asymmetrische Laplace-Verteilung hat dabei einen direkten Link zur Quantilsschätzung und wird genauer untersucht in Yu and Zhang [2005].
In linearen gemischten Modellen wurde die Quantilsschätzung unlängst von Geraci and Bottai [2007] und Geraci and Bottai [2014] entwickelt. Auch sie benutzen die Äquivalenz zu einem gemischten Modell mit asymmetrisch Laplace-verteilten Fehlern. Der Schätzung kommt die Verschiebung in einen Maximum Likelihood-Ansatz zu Gute. Ein numerischer Schätzalgorithmus ist in der Open Source-Software R implementiert (siehe das Paket lqmm von Geraci [2016]). Wegen der komplexen Darstellung der Log-Likelihood-Dichte sind analytische Lösungsansätze noch offen. Auch eine asymptotische Theorie wurde bis zu dieser Arbeit, welche die Konsistenz des bedingten Quantilsschätzers unter zusätzlichen Bedingungen zeigt, noch nicht entwickelt. Beim Beweis dieser Eigenschaft wird der Weiss'sche Satz (vgl. Weiss [1971] und Weiss [1973]) für abhängige Beobachtungen in der Maximum Likelihood-Schätzung angewandt. Im linearen gemischten Modell haben bereits Miller [1977] und Pinheiro [1994] diesen Satz benutzt, als sie die asymptotische Normalität der Parameterschätzung im Erwartungswertmodell bewiesen haben. In der Quantilsschätzung besteht eine Notwendigkeit für seine Anwendung darin, die zweiten Ableitungen der Log-Likelihood-Dichte bezüglich der unbekannten Parameter zu berechnen. Diese bilden eine Art der Fisher- Informations-Matrix. Deswegen bestimmen sie die asymptotische Varianz der Parameterschätzer und werden beim Beweis der Voraussetzungen des Weiss'schen Satzes benötigt. Die resultierende asymptotische Normalität der Parameterschätzer impliziert die Konsistenz eines bedingten τ-Quantils für ein gegebenen Wert $\tau \in (0, 1)$. Beide bewiesenen Eigenschaften, sowohl die asymptotische Normalität der Parameterschätzer als auch die Konsistenz des Quantilsschätzers, werden in modell-basierten Simulationsstudien bestätigt.
Die Anwendung der Quantilsregression im linearen gemischten Mo-

dell wird auf Zähldaten erweitert. Auch dafür wird die Konsistenz-Eigenschaft in dieser Arbeit bewiesen. Des Weiteren wird eine Methode, genannt Microsimulation via Quantiles, erarbeitet, mit welcher andere Populationsparameter als das arithmetische Mittel geschätzt werden können. Dabei wird der natürliche Zusammenhang zwischen Quantilen und der Verteilungsfunktion genutzt, welches in einer Schätzung der gesamten Verteilung mündet. Daraus kann nun jeder interessierende Parameter – z.B. ein Quantil, ein Anteil oder etwas anderes – über eine Monte Carlo- oder auch Mikrosimulation errechnet werden.

BIBLIOGRAPHY

J. Aldrich. R. A. Fisher and the Making of Maximum Likelihood 1912-1922. Statistical Science, 12(3):162–176, 1997.

H. Aman and J. Escher. Analysis II. Birkhäuser, Basel, 2006.

H. Aman and J. Escher. Analysis III. Birkhäuser, Basel, 2008.

G. W. Bassett and R. Koenker. Strong Consistency of Regression Quantiles and Related Empirical Processes. Econometric Theory, 2(2):191–201, 1986.

J. Breckling and R. Chambers. M-quantiles. Biometrika, 75 (4):761–771, 1988.

M. Buchinsky. Estimating the Asymptotic Covariance Matrix for Quantile Regression Models. A Monte Carlo Study. Journal of Econometrics, 68(2):303–338, 1995.

F. P. Cantelli. Sulla Determinazione Empirica della Legge di Probabilita. Giornale dell'Istituto Italiano degli Attuari, 4:221–424, 1933.

R. Chambers and N. Tzavidis. M-quantile models for small area estimation. Biometrika, 93 (2):255–268, 2006.

H. Cramér. Mathematical Methods of Statistics. Princeton University Press, Princeton, 1946.

B. Efron. Poisson Overdispersion Estimates on the Method of Asymmetric Maximum Likelihood. Journal of the American Statistical Association, 87(417):98–107, 1992.

M. Fox and H. Rubin. Admissibility of Quantile Estimates of a Single Location Parameter. The Annals of Mathematical Statistics, 35(3): 1019–1030, 1964.

H.-O. Georgii. Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik., volume 4. Auflage. Walter de Gruyter, 2009. doi: 10.1515/9783110215274.

M. Geraci. lqmm: Linear Quantile Mixed Models, 2016. URL https://cran.r-project.org/package=lqmm. R package version 1.5.3.

M. Geraci and M. Bottai. Quantile Regression for Longitudinal Data using the Asymmetric Laplace Distribution. Biostatistics, 8(1):140–154, 2007.

M. Geraci and M. Bottai. Linear Quantile Mixed Models. Statistics and Computing, 24(3):461–479, 2014.

V. Glivenko. Sulla Determinazione Empirica della Legge di Probabilita. Giornale dell'Istituto Italiano degli Attuari, 4:92–99, 1933.

X. He. Quantile Curves without Crossing. The American Statistician, 51(2):186–192, 1997.

C. R. Henderson. Estimation of Genetic Parameters. Annals of Mathematical Statistics, 21(2):309–310, 1950.

P. J. Huber. Robust Estimation of a Location Parameter. The Annals of Mathematical Statistical, 35:73–101, 1964.

P. J. Huber. Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In Berkeley: University of California Press, editor, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 221–233, 1967.

M. C. Jones. Expectiles and M-Quantiles are Quantiles. Statistics and Probability Letters, 20(2):149–153, 1994.

S. Jung. Quasi-Likelihood for Median Regression Models. Journal of the American Statistical Association, 91(a33):251–257, 1996.

A. Klenke. Wahrscheinlichkeitstheorie. Springer, Berlin, 2013.

R. Koenker. Quantile Regression for Longitudinal Data. Journal of Multivariate Analysis, 91(1):74–89, 2004.

R. Koenker. Quantile Regression. Cambridge University Press, Cambridge, 2005.

R. Koenker and G. Bassett. Regression Quantiles. Econometrica, 46 (1):33–50, 1978.

R. Koenker and J. A. F. Machado. Goodness of Fit and Related Inference Processes for Quantile Regression. Journal of the American Statistical Association, 94(448):1296–1310, 1999.

P.-S. Laplace. Mémoire sur la probabilié des causes par les évènements. Mémoires de l'Academie Royale des Sciences Presentés par Divers Savan, 6:621–656, 1774.

E. L. Lehmann and G. Casella. Theory of Point Estimation. Springer, New York, 1998.

J. A. F. Machado and J. M. C. Santos Silva. Quantiles for Counts. Journal of the American Statistical Association, 100(472):1226–1237, 2005.

J. R. Magnus. Maximum Likelihood Estimation of the GLS Model with Unknown Parameter in the Disturbance Covariance Matrix. Journal of Econometrics, 7(3):281–312, 1978.

A. A. Markov. Wahrscheinlichkeitsrechnung. Teubner, Leipzig, 1912.

W. J. McGill. Random fluctations of response rate. Psychometrika, 27: 3–17, 1962.

G. J. McLachlan and T. Krishnan. The EM Algorithm and Extensions. Wiley-Interscience, Hoboken, NJ, 2008.

J. J. Miller. Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance. The Annals of Statistics, 5(4):746–762, 1977.

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. Econometrica, 55 (4):819–847, 1987.

J. C. Pinheiro. Topics in Mixed Effects Models. PhD thesis, University of Wisconsin - Madison, 1994.

J. C. Pinheiro and E. C. Chao. Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models. Journal of Computational and Graphical Statistics, 15(1):58–81, 2006.

D. Pollard. Asymptotic for Least Absolute Deviation Regression Estimators. Econometric Theory, 7(2):186–199, 1991.

J. L. Powell. The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators. Econometrica, 51(5):1569–1575, 1983.

J. N. K. Rao. Small Area Estimation. Wiley, Hoboken, NJ, 2003.

G. K. Robinson. That BLUP is a Good Thing: The Estimation of Random Effects. Statistical Science, 6(1):15–32, 1991.

A. J. Rogers. Least Absolute Deviation Regression under Nonstandard Conditions. Econometric Theory, 17(4):820–852, 2001.

D. Ruppert, M.P. Wand, and R. Carroll. Semiparametric Regression. Cambridge University Press, Cambridge, 2003.

S. R. Searle. Linear Models. Wiley, New York, 1971.

S. R. Searle. Linear Models. Wiley, New York, 1997. Reprinted from 1971.

S. R. Searle, G. Casella, and C. E. McCulloch. Variance Components. Wiley, Hoboken, NJ, 1992.

R. Steyer. Wahrscheinlichkeit und Regression. Springer, Berlin, 2003.

G. Strang. Introduction to Linear Algebra, volume 5th Edition. Wellesley-Cambridge Press, 2016.

A. W. van der Vaart. Asymptotic Statistics. Cambridge University Press, Cambridge, 2007.

B. Weidenhammer, T. Schmid, N. Salvati, and N. Tzavidis. A Unit-level Quantile Nested Error Regression Model for Domain Prediction with Continuous and Discrete Outcomes. Working Papers Economics, School of Business & Economics, Freie Universität Berlin, 2016/12, 2016. URL http://edocs.fu-berlin.de/docs/receive/FUDOCS_document_000000024912.

L. Weiss. Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases. Journal of the American Statistical Association, 66(332):345–350, 1971.

L. Weiss. Asymptotic Properties of Maximum Likelihood Estimators in Some Nonstandard Cases, II. Journal of the American Statistical Association, 68(342):428–430, 1973.

K. Yu and R. Moyeed. Bayesian Quantile Regression. Statistics & Probability Letters, 54(4):437–447, 2001.

K. Yu and J. Zhang. A Three-Parameters Asymmetric Laplace Distribtuion and its Extension. Communication in Statistics - Theory and Methods, 34(9-10):1867–1879, 2005.

COLOPHON

The research leading to these results has received support under the European Commissions 7th Framework Programme (FP7/2013-2017) under grant agreement n312691, InGRID Inclusive Growth Research Infrastructure Diffusion.