



HAL
open science

Quantile regression for longitudinal data with controlled within-individual variance

Angelo Alcaraz, Gilles Durrieu, Audrey Poterie

► **To cite this version:**

Angelo Alcaraz, Gilles Durrieu, Audrey Poterie. Quantile regression for longitudinal data with controlled within-individual variance. 2024. hal-04542939

HAL Id: hal-04542939

<https://hal.science/hal-04542939>

Preprint submitted on 11 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantile regression for longitudinal data with controlled within-individual variance

Angelo Alcaraz, Gilles Durrieu and Audrey Poterie

*

Univ Bretagne Sud, CNRS UMR 6205, LMBA, F-56000 Vannes, France

April 11, 2024

Abstract

The presence of heteroscedasticity in data can often throw statistical modeling into disarray. In the context of mixed models and longitudinal data, this paper directly addresses this problem. We develop a quantile estimator based on the asymmetric Laplace distribution, which explains the heteroscedasticity between different groups of data. In addition to developing this new model, our paper establishes the good asymptotic properties of this estimator under minimal assumptions on the data and verifies them using simulations. Instead of improving performance point by point, our model focuses on the correct representation of data dispersion. Using the permissive formalism of the asymmetric Laplace distribution, we demonstrate the asymptotic properties of a class of estimators defined by a generalized optimization problem inspired by maximum likelihood. A Ridge penalization is proposed to address problems of variability overestimation. More generally, this paper presents a model for handling volume estimation problems more accurately. An application to the diet diversity of coral reef fish is proposed through the representation of isotopic niche sizes.

Keywords: Asymmetric Laplace Distribution, Linear Quantile Mixed Models, Gaussian Quadrature, Ridge Regression, Penalized Linear Quantile Mixed Models

*The authors gratefully acknowledge ANR TONIC.

1 Introduction

In this article, we propose a formal theory to deal with heteroscedasticity in a panel data model. In certain applications, it may be useful to have a precise estimator of data dispersion, in particular to measure heteroscedasticity. This is what the model in this article attempts to achieve, by explicitly mentioning the differences in variability between different groups of data. What's more, the underlying mathematical formalism proposed allows for the simple addition of penalties, thus allowing control on the modelization error. By studying several independent variables of interest using the model in this paper, it is then possible to perform volume estimation in the broad sense (for spaces of dimensions greater than 1), controlling the within-group dispersion of each of these variables.

Here, we consider quantile regression models to provide robustness against outliers. Koenker and Basset introduced quantile regression in 1978 (Koenker and Bassett (1978)) as a more robust alternative to classical linear least-squares regression. An overview of robust statistics literature can be found in Hampel et al. (2011), Jurecková and Sen (1996), Hubert and Branden (2003) and Koenker et al. (2017). Instead of focusing on the changes in the mean, the quantile regression approach allows one to test whether there is a change in the τ th quantile of the model response for any given $\tau \in (0, 1)$. When the conditional distributions are non-Gaussian, the mean might not be the best summary, and a change in distributions may not be detected. Inference for linear quantile regression models has become a subject of intense investigation in the past years. Our article establishes a general theoretical framework based on regression quantile for panel data allowing for heteroscedasticity in the data. We define a new class of estimators by adding a term to the optimization problem defining the maximum likelihood estimator. One of the main advantages of our method is the generalization of good asymptotic properties to this class of

estimators. This makes it possible to implement a wide choice of penalty, introducing a bias which could compensate for the modeling error. A penalization is also introduced to approach between-measurements variability in a better way. Instead of showing better performances, this new modelization aims at making the proposed model more explicable in order to predict and apprehend multi-level variability more effectively.

The development of quantile regression is linked to a desire to get away from the “normality dogma”, following several criticisms of the Gaussian distribution, notably its absence in certain econometric applications (Stigler (1973)). More precisely, instead of considering the calculated response as an average using least-squares minimization, we use a minimization problem defining the quantiles of a series of observations to establish an estimator. This provides greater resistance to the addition of outliers and relaxes assumptions on the residuals.

We take the following linear regression model defined by :

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n \quad (1)$$

where for all $n \geq 1$, $\mathbf{Y}_n = (Y_1, \dots, Y_n)^t$ is the vector of observations, \mathbf{X}_n is a known matrix of dimension $n \times p$ with rows $\mathbf{x}_i^t \in \mathbb{R}^p$, $i = 1, \dots, n$, $\boldsymbol{\epsilon}_n = (\epsilon_1, \dots, \epsilon_n)^t$ is a vector of independent and identically distributed (i.i.d.) errors and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ denotes the unknown vector of regression parameters to be estimated.

We call τ -regression quantile any solution of the minimization problem

$$\boldsymbol{\beta}(\tau) = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(Y_i - \mathbf{x}_i^t \boldsymbol{\beta}) \quad (2)$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}_{u < 0})$ with $\mathbb{1}_{\mathcal{P}}$ takes the value 1 or 0 depending on whether the condition \mathcal{P} is satisfied or not. The function ρ is called the “quantile loss function” and is classically used to define the quantile of a random variable. This estimator shares the asymptotic properties of a

classical least-square regression estimator, with less restrictive assumptions (Koenker and Bassett (1978)).

This specificity allows a wide range of uses of the quantile regression models. We can cite applications to sociological data (Abrevaya and Dahl (2008)), medical data (Peng et al. (2009)), genomic data (Durrieu and Briollais (2009) and Briollais and Durrieu (2017)) or ecological data (Cade and Noon (2003)).

The development of quantile regression for panel data kind of follows the path of the development of generalized linear models (Liang and Zeger (1986)). Koenker (2004) was interested in longitudinal data, a classical form of data that can be found in medicine, ecology, or biology. In this paradigm, we account for dependency between grouped measurements (for example a medical measurement on the same subject). This led to the formulation of a quantile regression model with fixed effects, expressed as a lasso (least absolute shrinkage and selection operator) penalty (Lamarche (2010)).

The link with generalized models only got stronger, notably with an approach to quantile regression with random effects (Arellano and Bonhomme (2013), Galvao and Poirier (2019)). Several quantile regression-based modeling approaches were introduced, such as Bayesian quantile modeling (Alhamzawi and Yu (2014), Ji and Shi (2022)), semi-parametric modeling (Kim and Yang (2011)) or composite modeling (Wang and Xiang (2017), Wu and Yao (2016), Lu and Fan (2015)). A summary of advances in quantile regression and its applications is available in Yu et al. (2003) and Koenker (2017).

In this article, we consider the development of quantile modeling based on the asymmetric Laplace distribution (Yu and Zhang (2005)). Several seminal papers have established the link between the quantile regression minimization problem and the maximum likelihood estimator (Yu and Moyeed (2001), Geraci and Bottai (2007)). This made it possible to use classical likelihood

estimator resolution algorithms to determine the quantile regression estimator. Note the use of Bayesian theory (Luo et al. (2012), Alhamzawi and Yu (2014), Aghamohammadi and Mohammadi (2017), Ji and Shi (2022)) or the *Expectation-Maximization* algorithm (Geraci and Bottai (2014), Geraci (2014), Galarza et al. (2017) Tian et al. (2020), Battagliola et al. (2022)). This article follows in these authors' footsteps, exploiting the asymmetric Laplace distribution and its link with quantile regression, using it on longitudinal data with random effects.

Section 2 is dedicated to the description of the model, its properties, and the numerical estimation. Our model is first introduced. Our approach makes it possible to include heteroscedasticity in the model explicitly. Next, the asymptotic statistical properties of the proposed quantile estimator are studied. A generalization of our model is also introduced, along with its asymptotic properties. Using this generalization, we can deal with the overestimation of dispersion. The sketch of the proofs is given in the same Section 2. Next, our estimation procedures are explained. Next, the asymptotic properties of our approach are illustrated in Section 3 on simulated data and an application of the method on ecological data is provided in Section 4. The latter highlights the value of our method in terms of surface estimation. This application was crucial to our methodological development, as it was the original motivation for developing our quantile estimator. Finally, Section 5 concludes with a general discussion of our approach and its feasibility and applicability in practice. All the proofs are postponed in the supplementary material. The estimation algorithms can be found at <https://github.com/I621974/hlqmm>.

2 Model and estimation

2.1 Quantile regression for heteroscedastic longitudinal data using the asymmetric Laplace distribution

We consider the framework of longitudinal data. Let $n \in \mathbb{N}$ be the number of individuals. The index $i \in \{1, \dots, n\}$ corresponds to the individual levels and for each individual i , there are $n_i \in \mathbb{N}$ measurements. The index $j \in \{1, \dots, n_i\}$ corresponds to the j -th measure of the i -th individual. Let $\mathbf{Y} = (Y_{ij})_{1 \leq i \leq n, 1 \leq j \leq n_i} \in \mathbb{R}^N$ be a response variable, with $N = \sum_{i=1}^n n_i$.

The general form of the mixed model is

$$Y_{ij} = \mathbf{X}_{ij}^t \boldsymbol{\beta} + \mathbf{e}_i^t \boldsymbol{\nu} + \epsilon_{ij} \quad (3)$$

where for all $i \in \{1, \dots, n\}$ and all $j \in \{1, \dots, n_i\}$, $\mathbf{X}_{ij} \in \mathbb{R}^p$ are the observations for the j -th measure of the i -th individual, $p \in \mathbb{N}$ is the number of variables, $\boldsymbol{\beta} \in \mathbb{R}^p$ are the regression parameters (who can contain fixed effects), $\mathbf{e}_i \in \mathbb{R}^n$ is the i -th vector of the standard basis, $\boldsymbol{\nu} \in \mathbb{R}^n$ is the vector of random effects, describing a individual-level effect as an intercept, and ϵ_{ij} is a centered random variable representing the error term.

Following Geraci and Bottai (2014), we have two more assumptions on the model. First, let $\boldsymbol{\nu} \sim \mathcal{N}(0, \phi \boldsymbol{\Sigma})$, with $\phi \in \mathbb{R}$ denoting a fixed parameter and $\boldsymbol{\Sigma} \in \mathbb{R}^{n^2}$ being a known positive definite matrix. The latter is given later and shall represent the dependency structure between individuals. Secondly, we suppose that ϵ_{ij} follows an asymmetrical Laplace distribution, similarly to Geraci and Bottai (2007). We say that a random variable Z follows an asymmetrical Laplace distribution of parameters (μ, σ, τ) (which is denoted by $Z \sim ALD(\mu, \sigma, \tau)$) if its density can be expressed as:

$$f_Z(z) = \frac{\tau(1-\tau)}{\sigma} \exp \left(-\rho_\tau \left(\frac{z-\mu}{\sigma} \right) \right)$$

with $\rho_\tau(u) = u(\tau - \mathbb{1}_{u < 0})$ the quantile loss function. For more information about the asymmetric Laplace distribution, see Yu and Zhang (2005). More precisely, we assume that $\epsilon_{ij} \sim ALD(0, \sigma_i, \tau)$, where $0 < \tau < 1$ denotes the order of the quantile and $\boldsymbol{\sigma} = (\sigma_i)_{1 \leq i \leq n} \in (\mathbb{R}_+^*)^n$ corresponds to the scale parameters. Using this assumption, we explicitly take into account the heteroscedasticity that exists between individuals. Moreover, we suppose that all ϵ_{ij} are independent from each other and from the random effects. Finally, we have to infer the vector $\boldsymbol{\theta} = (\boldsymbol{\sigma}, \boldsymbol{\beta}) \in (\mathbb{R}_+^*)^n \times \mathbb{R}^p$.

This framework is inspired by Geraci (2014) but, in this former article, they assumed that for all $i \in \{1, \dots, n\}, \sigma_i = \sigma \in \mathbb{R}_+^*$. What we are doing here is allowing some heteroscedasticity between individuals by letting the scale parameters depend on i . This is crucial to our work as we are interested in controlling within-individual variations in the presence of heteroscedasticity for longitudinal data. This is a new approach and one of the main differences with previous quantile estimators for mixed models.

In an inference framework, it is instructive to study the asymptotic properties of an estimator. Asymptotic normality is of particular interest, as it makes it easy to control the estimator's behavior at infinity. These properties have already been demonstrated in mixed models for least-squares estimators (Demidenko (1997), Jiang (1998), Jiang (2017)) and other quantile estimators (Sheather and Marron (1990), Koenker (1994), Koenker (2005), Koenker et al. (2017)).

Regarding asymptotics, one consequence of this modeling choice is the way to tend toward infinity. As the number of parameters to estimate now depends on n , we are not interested in the limit when n tends to infinity. Instead, we determine the limit of our estimator when for all $1 \leq i \leq n, n_i \rightarrow +\infty$, that is when the number of measures per individual tends to infinity.

We define our estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ as a maximum likelihood estimator of $\boldsymbol{\theta}$. Let

$$f_{Y_{ij}|\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\sigma}}(y_{ij}, \mathbf{v}) = \frac{\tau(1-\tau)}{\sigma_i} \exp\left(-\rho_\tau\left(\frac{y_{ij}-\mu_{ij}}{\sigma_i}\right)\right)$$

be the density of the response conditionally on the parameters with $\mu_{ij} = \mathbf{X}_{ij}^t \boldsymbol{\beta} + \mathbf{e}_i^t \mathbf{v}$. Because of the independence assumptions, we obtain for all \mathbf{y} and \mathbf{v} the multivariate density:

$$f_{\mathbf{Y},\boldsymbol{\nu}|\boldsymbol{\theta}}(\mathbf{y}, \mathbf{v}) = \prod_{i=1}^n \prod_{j=1}^{n_i} f_{Y_{ij}|\boldsymbol{\beta},\boldsymbol{\nu},\boldsymbol{\sigma}}(y_{ij}, \mathbf{v}) f_{\boldsymbol{\nu}|\boldsymbol{\phi}}(\mathbf{v}). \quad (4)$$

By integrating (4) on the random effects, we obtain the likelihood function. We first have:

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^n} f_{\mathbf{Y},\boldsymbol{\nu}|\boldsymbol{\theta}}(\mathbf{y}, \mathbf{v}) d\mathbf{v} \quad (5)$$

and the estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \left(\log \left(\int_{\mathbb{R}^n} f_{\mathbf{Y},\boldsymbol{\nu}|\boldsymbol{\theta}}(\mathbf{y}, \mathbf{v}) d\mathbf{v} \right) \right) \quad (6)$$

with $\Theta \subset (\mathbb{R}_+^*)^n \times \mathbb{R}^p$ the parameter space.

The minimization problem is equivalent to the quantile regression estimator minimization problem (Koenker and Bassett (1978)). This equivalence is the main motivation behind the use of the asymmetric Laplace distribution.

Furthermore, we define an additional estimator $\hat{\boldsymbol{\theta}}_f$ of $\boldsymbol{\theta}$ which is defined as the solution of the following maximization problem

$$\hat{\boldsymbol{\theta}}_f = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R}^n} \left(\sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left(\frac{y_{ij}-\mu_{ij}}{\sigma_i} \right) + \sum_{i=1}^n n_i \ln(\sigma_i) + \sum_{i=1}^n f_{n_i}(\boldsymbol{\theta}) \right) d\mathbf{v} \quad (7)$$

where $(f_{n_i})_{1 \leq i \leq n}$ is a sequence of functions taking its values in the space of parameter Θ . This optimization problem corresponds to the maximum likelihood optimization problem with the penalty term $\sum_{i=1}^n f_{n_i}(\boldsymbol{\theta})$ added. Such regularization term can be added to the maximum likelihood optimization problem to reduce model complexity and prevent over-fitting (Tibshirani

(1996), Friedman et al. (2000)). Moreover, one specificity of this modelization is the possibility to easily consider a modified optimization problem by only multiplying the data by an appropriate quantity, as in equation (7).

As quantile regression grows more complex, for instance with the addition of hierarchy or random effects, resolving the original optimization problem defining the quantile regression can be challenging. No exact solution exists and then using some numerical algorithms to get a solution can easily be computationally expensive (Koenker (2005), Koenker et al. (2017)). Nonetheless, the asymmetric Laplace distribution allows us to use all the algorithms available for maximum likelihood estimation (like the EM algorithm) to compute the quantile estimator (Booth and Hobert (1999), Geraci and Bottai (2007), Galarza et al. (2017)).

Most of the computational problems are shared by the method of Geraci and Bottai (2007) and Geraci and Bottai (2014) and developed in detail in these articles. The computation time necessary for the estimation can remain quite long and the optimization method used will suffer from a curse of dimensionality when n grows (and therefore the number of parameters). Where our model gains in explainability by modeling the parameters $\sigma_i, 1 \leq i \leq n$, it loses in computational ease.

2.2 Theoretical results

In order to investigate the asymptotic behavior of our estimates, it is necessary to introduce several assumptions.

A1. We have $N \rightarrow +\infty$ and for all $i \in \{1, \dots, n\}$, $n_i \rightarrow +\infty$ so that $n_i = \mathcal{O}(N)$.

A2. The true value of θ is denoted by θ^0 and we have $\theta^0 \in \text{int}(\Theta)$, where $\text{int}(\Theta)$ is the interior of Θ .

A3. For all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n_i\}$, the random variables ϵ_{ij} are independent from each other and independent from ν .

A4. The sequence $\frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_{ij}$ tends to $\mathbf{c}_i \in \mathbb{R}^p$ when $n_i \rightarrow +\infty$ for all $i \in \{1, \dots, n\}$.

A5. The sequence $\frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{X}_{ij} \mathbf{X}_{ij}^t$ tends to a positive definite matrix $\mathbf{C}_i \in \mathbb{R}^{p \times p}$ when $n_i \rightarrow +\infty$ for all $i \in \{1, \dots, n\}$.

A6. For all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, n_i\}$, the random variables $\mathbb{E}[\epsilon_{ij} | \mathbf{Y}]$ are independent from each other and independent from $\mathbb{E}[\nu | \mathbf{Y}]$.

Assumption (A1) is linked to the way we reach infinity as explained in Section 2.1. The particularity here is that n stays constant but N reaches infinity because all the $n_i, 1 \leq i \leq n$ go to infinity. Assumption (A2) is a classical assumption for quantile regression and is needed for the main theorem used during the proof of the asymptotic behavior of our estimator. Assumptions (A3) to (A5) constrain the form the data can take and ensure the existence of first- and second-order moments. These assumptions are also commonly used in quantile regression. We can see that neither the data nor the variability term ϵ_{ij} need to follow a normal distribution for asymptotic results to be established. The last assumption (A6) is needed for the proof and is rarely used. We can see an usage and a detailed explanation of this assumption in Weidenhammer (2017).

The main results of the paper is the following theorem which provides the asymptotic normality of our estimates.

Theorem 1. *Assume assumptions (A1) to (A6) hold. Then we have as N tends to infinity the asymptotic normality*

$$\text{diag}((\sqrt{N}, \sqrt{N})) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \begin{pmatrix} \sqrt{N}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}^0) \\ \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, B^{-1}(\boldsymbol{\theta}^0)) \quad (8)$$

where $B(\boldsymbol{\theta}^0)$ is a definite positive $(n+p) \times (n+p)$ matrix.

Moreover, we can compute the value of $B(\boldsymbol{\theta}^0)$, allowing us to establish precise confidence intervals. This also enables us to compare the asymptotic distribution of the parameter with the estimator one and gives us information about the rate of convergence, that is to say, \sqrt{N} . The consistency of our estimator will immediately follow this result, as in Weidenhammer (2017).

We define two more assumptions on $(f_{n_i})_{1 \leq i \leq n}$ from equation (7):

C1. For all $1 \leq i \leq n$, we have $\frac{f_{n_i}}{n_i} \rightarrow l \in \mathbb{R}$ when n_i tends to infinity.

C2. The application

$$\begin{aligned} \Phi : \Theta &\longrightarrow (\mathbb{R}_+^*)^n \\ \boldsymbol{\theta} &\longrightarrow \left(\sigma_i \exp\left(\frac{f_{n_i}(\boldsymbol{\theta})}{n_i}\right) \right)_{1 \leq i \leq n} \end{aligned}$$

is a bijective function.

Theorem 2. Assume assumptions (A1) to (A6), (C1) and (C2) hold. Then, we have as N tends to infinity the asymptotic normality

$$\text{diag}((\sqrt{N}, \sqrt{N})) (\hat{\boldsymbol{\theta}}_f - \tilde{\boldsymbol{\theta}}^0) = \begin{pmatrix} \sqrt{N}(\hat{\boldsymbol{\sigma}}_f - \tilde{\boldsymbol{\sigma}}^0) \\ \sqrt{N}(\hat{\boldsymbol{\beta}}_f - \boldsymbol{\beta}^0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{B}^{-1}(\boldsymbol{\theta}^0)) \quad (9)$$

where $\tilde{\boldsymbol{\theta}}^0 = (\tilde{\boldsymbol{\sigma}}^0, \boldsymbol{\beta}^0)$, for all $i \leq n$, $\tilde{\sigma}_i^0 = \exp\left(\frac{f_{n_i}(\boldsymbol{\theta}^0)}{n_i}\right)\sigma_i^0$ and

- $\tilde{B}_{\boldsymbol{\sigma}, \boldsymbol{\sigma}}(\boldsymbol{\theta}^0) = e^{2l} B_{\boldsymbol{\sigma}, \boldsymbol{\sigma}}(\boldsymbol{\theta}^0)$,
- $\tilde{B}_{\boldsymbol{\sigma}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0) = e^l B_{\boldsymbol{\sigma}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0)$,
- $\tilde{B}_{\boldsymbol{\beta}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0) = B_{\boldsymbol{\beta}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0)$.

Theorem 2 will allow us to ensure asymptotic normality for an all-new class of estimators $\boldsymbol{\theta}_f$.

Therefore, by using an appropriate sequence of functions $(f_{n_i})_{1 \leq i \leq n}$, we can quite modify the

optimization problem (7) while keeping the same theoretical guarantees. Here, we are interested in the case where the function f is as ridge penalization and so we want to solve

$$\hat{\boldsymbol{\theta}}_f = \arg \max_{\boldsymbol{\theta} \in \Theta} \int_{\mathbb{R}^n} \left(\sum_{i=1}^n \sum_{j=1}^{n_i} \rho_{\tau} \left(\frac{y_{ij} - \mu_{ij}}{\sigma_i} \right) + \sum_{i=1}^n n_i \ln(\sigma_i) + \lambda \sum_{i=1}^n \sigma_i^2 \right) d\mathbf{v} \quad (10)$$

We can apply Theorem 2 to (10) in which $f_{n_i}(\boldsymbol{\theta}) = \lambda \sigma_i^2$, for all $1 \leq i \leq n$ and with $\lambda > 0$ a fixed parameter. We have the condition (C1) as, for all $1 \leq i \leq n$, we have $\frac{f_{n_i}}{n_i} \rightarrow 0$.

In this case, the assumption (C2) also holds. Indeed, if we write $\tilde{\sigma}_i = \sigma_i \exp\left(\lambda \frac{\sigma_i^2}{n_i}\right)$, we can obtain the desired scale parameters σ_i in the following way:

$$\sigma_i = \frac{\sqrt{W\left(\frac{2\lambda\tilde{\sigma}_i^2}{n_i}\right)} n_i}{\sqrt{2\lambda}}$$

with W being the Lambert function, defined as the inverse function of $x \rightarrow xe^x$.

As a consequence, the estimator defined by the optimization problem (7) including a Ridge penalization term $\lambda \sum_{i=1}^n \sigma_i^2$ is asymptotically normal, with the same variance-covariance matrix as the non-penalized problem (as $l = 0$ here). We can hope that the ridge penalization allows for greater control over the estimation of the scale parameters of the model. Moreover, if $Z \sim ALD(\mu, \sigma, \tau)$, we have the following result (Yu and Zhang (2005)):

$$Var(Z) = \frac{\sigma^2(1 - 2\tau + 2\tau^2)}{(1 - \tau)^2\tau^2}. \quad (11)$$

As σ_i is the scale parameter for the i -th individual, one can observe that penalizing σ_i allows us to deal with over-estimation of the variance between measures of the same individual. In some way, information about the scale parameter σ_i gives us knowledge about the dispersion or diversity of measures for an individual.

We want to note that model error and measurement error aren't taken into account in our linear mixed model. Indeed, the between-measurements variability is considered as a consequence of natural differences between measurements. Therefore, we hope that controlling λ allows us to

cope with the differences between our model and real data. Adding a Ridge penalization to our model will add bias to our estimates, but our aim is to accurately represent the dispersion, not necessarily to have a better point-by-point estimate.

The next section provides the proofs of Theorem 1 and Theorem 2.

2.3 Proofs

Proof of Theorem 1. We follow the same steps proposed in Weidenhammer (2017). To establish the asymptotic normality, we shall make use of the Theorem 3 of Weiss (Weiss (1971, 1973)):

Theorem 3. *Let N observations Y_1, \dots, Y_N from a known distribution $P_{Y|\theta}$ with $\theta \in \Theta \subset \mathbb{R}^k$ a parameter vector in the parameter space of dimension k . Considering the density $f_{\mathbf{Y}}(\mathbf{y}|\theta)$ of the vector $\mathbf{Y} = (Y_1, \dots, Y_N)$, we can define the log-likelihood $l(\theta|\mathbf{Y}) = \log f_{\mathbf{Y}}(\mathbf{y}|\theta)$. Let $\theta^0 \in \text{int}(\Theta)$ be the true value of the parameter vector and assume there exist $2k$ sequences $K_1(N), \dots, K_k(N)$ and $M_1(N), \dots, M_k(N)$ such that, for all $\iota \in \{1, \dots, k\}$, $K_\iota(N) \xrightarrow{N \rightarrow +\infty} +\infty$, $M_\iota(N) \xrightarrow{N \rightarrow +\infty} +\infty$ and $\frac{M_\iota(N)}{K_\iota(N)} \xrightarrow{N \rightarrow +\infty} 0$.*

We assume the two following conditions:

B1. We have the following convergence in probability when N tends to infinity:

$$-\frac{1}{K_{\iota_1}(N)K_{\iota_2}(N)} \frac{\partial^2}{\partial \theta_{\iota_1} \partial \theta_{\iota_2}} l(\theta|\mathbf{Y}) \Big|_{\theta=\theta^0} \xrightarrow{P} B_{\iota_1, \iota_2}(\theta^0)$$

where $B_{\iota_1, \iota_2}(\theta^0)$ is continuous and the $k \times k$ matrix $B(\theta^0)$ is positive definite.

B2. There exists two positive sequences $(\gamma(N, \theta_0))_N$ and $(\delta(N, \theta_0))_N$ so that $(\gamma(N, \theta_0))_N \xrightarrow{N \rightarrow +\infty} 0$ and $(\delta(N, \theta_0))_N \xrightarrow{N \rightarrow +\infty} 0$ and we have for all $\theta \in N_n(\theta^0)$

$$\mathbb{P}_\theta(R_n(\theta^0, \gamma(N, \theta_0))) > 1 - \delta(N, \theta_0)$$

with

- $N_n(\boldsymbol{\theta}^0) = \left\{ (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k)^t ; |\boldsymbol{\theta}_\iota - \boldsymbol{\theta}_\iota^0| \leq \frac{M_\iota(N)}{K_\iota(N)}, \iota \in \{1, \dots, k\} \right\}$,
- $R_n(\boldsymbol{\theta}^0, \gamma) = \left\{ \mathbf{Y} \in \mathbb{R}^N ; \sum_{\iota_1, \iota_2=1}^k M_{\iota_1}(N)M_{\iota_2}(N) \sup_{\boldsymbol{\theta} \in N_n(\boldsymbol{\theta}^0)} |\varepsilon_{\iota_1, \iota_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^0, n, \mathbf{Y})| \leq \gamma \right\}$,
- $\varepsilon_{\iota_1, \iota_2}(\boldsymbol{\theta}, \boldsymbol{\theta}^0, n, \mathbf{Y}) = -\frac{1}{K_{\iota_1}(N)K_{\iota_2}(N)} \frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) - B_{\iota_1, \iota_2}(\boldsymbol{\theta}^0)$.

Under all these assumptions, we have the existence of a maximum likelihood estimator sequence

$\hat{\boldsymbol{\theta}}(N)$ of $\boldsymbol{\theta}$, roots of the equation $\frac{\partial l(\boldsymbol{\theta} | \mathbf{Y})}{\partial \boldsymbol{\theta}} = 0$, so that:

$$\text{diag}(K_1(N), \dots, K_k(N)) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B^{-1}(\boldsymbol{\theta}^0)). \quad \square$$

In order to apply Theorem 3, we shall check the conditions **B1** and **B2**.

Checking of the Condition **B1**

The condition **B1** asserts the convergence of the second derivatives of the log-likelihood. From the weak law of large numbers, it is possible to relate these derivatives to their expectations. We can find a more precise description of this assumption and its relation to the Fisher information in Weidenhammer (2017), page 70 – 72.

Considering the Theorem of Schwartz, we only have to calculate the second derivatives of the log-likelihood with respect to $(\boldsymbol{\sigma}, \boldsymbol{\sigma})$, $(\boldsymbol{\sigma}, \boldsymbol{\beta})$ and $(\boldsymbol{\beta}, \boldsymbol{\beta})$.

One can find the calculation of the second derivatives in our case in the supplementary material in Section 6. We can notice that every second derivative can be expressed as a function of conditional derivatives with respect to \mathbf{Y} . It follows from Weidenhammer (2017), pages 81-82 that these conditional expectations can be expressed as an integral with respect to the measure $P_{\nu | \mathbf{Y}}$.

After calculating all the second derivatives, we obtain from the weak law of large numbers the expression of $B_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(\boldsymbol{\theta}^0)$ for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \{\boldsymbol{\sigma}, \boldsymbol{\beta}\}$. The sequences $K_\iota(N)$, for all ι are exactly the rate of convergence of the corresponding law of large numbers. Let ξ, ξ_1, ξ_2 3 random variables, it follows from Steyer (2013) :

$$\mathbb{E} [\text{Var}(\xi | \mathbf{Y})] = \text{Var}(\xi) - \text{Var}(\mathbb{E}[\xi | \mathbf{Y}])$$

and

$$\mathbb{E} [\text{Cov} (\xi_1, \xi_2 | \mathbf{Y})] = \text{Cov} (\xi_1, \xi_2) - \text{Cov} (\mathbb{E} [\xi_1 | \mathbf{Y}], \mathbb{E} [\xi_2 | \mathbf{Y}]).$$

Details on the calculation of $B_{\theta_1, \theta_2}(\boldsymbol{\theta}^0)$ for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \{\boldsymbol{\sigma}, \boldsymbol{\beta}\}$ are in the supplementary material in Section 6. We obtain from (20) and (21)

$$(K_{\boldsymbol{\sigma}}(N), K_{\boldsymbol{\beta}}(N)) = (\sqrt{N}, \sqrt{N}), \quad B_{\boldsymbol{\sigma}, \boldsymbol{\sigma}}(\boldsymbol{\theta}^0) = (b_{\boldsymbol{\sigma}, \boldsymbol{\sigma}}^{i,k})_{1 \leq i, k \leq n}$$

where

$$b_{\boldsymbol{\sigma}, \boldsymbol{\sigma}}^{i,k} = \frac{\mathbb{E} [\mathbb{E}^2 [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]] |_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \mathbb{1}_{k=i}}{(\sigma_i^0)^4}.$$

Furthermore, it follows from (22)

$$B_{\boldsymbol{\sigma}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0) = B_{\boldsymbol{\beta}, \boldsymbol{\sigma}}(\boldsymbol{\theta}^0) = (b_{\boldsymbol{\sigma}, \boldsymbol{\beta}}^{i,k})_{1 \leq i \leq n, 1 \leq k \leq p}$$

with

$$b_{\boldsymbol{\sigma}, \boldsymbol{\beta}}^{i,k} = \frac{\mathbf{c}_{ik}}{(\sigma_i^0)^2} \left(\tau + \frac{1}{\sigma_i^0} \mathbb{E} [\mathbb{E} [\epsilon_{i1} (\mathbb{1}_{\epsilon_{i1} \leq 0} - \tau) | \mathbf{Y}]] \mathbb{E} [\mathbb{1}_{\epsilon_{i1} \leq 0} | \mathbf{Y}] |_{\boldsymbol{\theta} = \boldsymbol{\theta}^0} \right),$$

and

$$B_{\boldsymbol{\beta}, \boldsymbol{\beta}}(\boldsymbol{\theta}^0) = \sum_{i=1}^n \frac{\mathbb{E} [\mathbb{E}^2 [\mathbb{1}_{\epsilon_{i1} \leq 0} | \mathbf{Y}]] - \tau}{(\sigma_i^0)^2} \mathbf{C}_i.$$

For all i such that $\sigma_i^0 \neq 0$ (always true by definition), the continuity of $B(\boldsymbol{\theta}^0)$ is obtained by the composition of continuous function.

We have the positive definiteness of $B(\boldsymbol{\theta}^0)$ in the same way as in Weidenhammer (2017), page 110 – 112. The form of the inverse of the variance-covariance matrix is very close to the one developed in this thesis, composed mainly of expectation or expectation of conditional expectation with respect to \mathbf{Y} . However, two main differences in our article are that the number of individuals n is not set to $+\infty$ and, as explained before, the $(\sigma_i)_{1 \leq i \leq n}$ are not equal, allowing for heteroscedasticity. This also changes the asymptotic properties of our estimates.

Checking of the Condition **B2**

Condition **B2** determines the behavior of the second derivatives of the likelihood at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ when N tends to infinity. More precisely, this condition allows the development of a Taylor series around the true parameter vector $\boldsymbol{\theta}^0$. Further details on this condition can also be found in Weidenhammer (2017), page 70 – 72. In the same way as in Miller (1977) and Pinheiro (1994), we define the following quantity:

$$\kappa(N) = \max_{\iota_1, \iota_2} \left| -\frac{1}{K_{\iota_1}(N)K_{\iota_2}(N)} \mathbb{E}_{\boldsymbol{\theta}^0} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] - B_{\iota_1, \iota_2}(\boldsymbol{\theta}^0) \right|$$

which allows us to construct for all ι the following sequences:

$$M \equiv M_\iota(N) = \min \left\{ \sqrt{N}^{\frac{1}{4}}, \kappa^{-\frac{1}{4}}(N) \right\}.$$

The value of M is chosen so as to make future proofs of convergence possible. In order to prove condition **B2**, it is only necessary to check that for ι_1, ι_2 and for all $\boldsymbol{\theta}^2 \in N_N(\boldsymbol{\theta}^0)$

$$M^2 \sup_{\boldsymbol{\theta}^1 \in N_n(\boldsymbol{\theta}^0)} \left| -\frac{1}{K_{\iota_1}(N)K_{\iota_2}(N)} \frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^1} - B_{\iota_1, \iota_2}(\boldsymbol{\theta}^0) \right| \xrightarrow{P_{\mathbf{Y}|\boldsymbol{\theta}^2}} 0. \quad (12)$$

This difference in (12) can be decomposed in the following way:

$$\begin{aligned} & \sup_{\boldsymbol{\theta}^1 \in N_n(\boldsymbol{\theta}^0)} \left(-\frac{1}{K_{\iota_1} K_{\iota_2}} \frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^1} - B_{\iota_1, \iota_2}(\boldsymbol{\theta}^0) \right) \\ &= \sup_{\boldsymbol{\theta}^1 \in N_n(\boldsymbol{\theta}^0)} \underbrace{\left(-\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^1} - \frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^2} \right)}_{=\Phi_1} \right) \\ & \quad - \underbrace{\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^2} - \mathbb{E}_{\boldsymbol{\theta}^2} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^2} \right]} \right)}_{=\Phi_2} \\ & \quad - \underbrace{\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\mathbb{E}_{\boldsymbol{\theta}^2} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^2} \right] - \mathbb{E}_{\boldsymbol{\theta}^2} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right]} \right)}_{=\Phi_3} \\ & \quad - \underbrace{\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\mathbb{E}_{\boldsymbol{\theta}^2} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] - \mathbb{E}_{\boldsymbol{\theta}^0} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right]} \right)}_{=\Phi_4} \\ & \quad - \underbrace{\frac{1}{K_{\iota_1} K_{\iota_2}} \left(\mathbb{E}_{\boldsymbol{\theta}^0} \left[\frac{\partial^2}{\partial \boldsymbol{\theta}_{\iota_1} \partial \boldsymbol{\theta}_{\iota_2}} l(\boldsymbol{\theta} | \mathbf{Y}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] - B_{\iota_1, \iota_2}(\boldsymbol{\theta}^0) \right)}_{=\Phi_5}. \end{aligned}$$

By a careful analysis of each term in the decomposition, by definition of M , we have

$$|\Phi_5| \leq \kappa \implies M^2 \Phi_5 \xrightarrow{N \rightarrow +\infty} 0.$$

To prove the convergence for the other part of this equation, we mainly use the fact that $\theta^1, \theta^2 \in N_N(\theta^0)$, knowing that we can control the rate of convergence between these parameters and θ^0 when $N \rightarrow +\infty$. The computation of Φ_1, Φ_2, Φ_3 and Φ_4 is quite tedious and so is left as supplementary material in Section 6. We can however notice that Φ_2 is a difference between a random variable and its expectation, which allows the use of the Bienaymé-Tchebychev inequality to show the appropriate convergence.

Once this is done, we are in the conditions of application of the Theorem 3 of Weiss under the assumptions (A1) to (A6). Weidenhammer (2017) page 146-147 showed that the asymptotic normality implied the consistency of the estimator $\hat{\theta}$ of θ . From the Theorem 3 of Weiss, we obtain successfully the asymptotic normality of our estimator. \square

Proof of Theorem 2. We consider the expanded data \tilde{Y} and \tilde{X} so that $\tilde{Y}_{ij} = \exp\left(\frac{f_{n_i}(\theta)}{n_i}\right) Y_{ij}$ and $\tilde{X}_{ij} = \exp\left(\frac{f_{n_i}(\theta)}{n_i}\right) X_{ij}$. We have the following linear mixed model:

$$\tilde{Y}_{ij} = \tilde{X}_{ij}^t \beta + \exp\left(\frac{f_{n_i}(\theta)}{n_i}\right) e_i^t \nu + \tilde{\epsilon}_{ij} \quad (13)$$

with $\tilde{\epsilon}_{ij} \sim ALD(0, \sigma_i \exp\left(\frac{f_{n_i}(\theta)}{n_i}\right), \tau)$.

As assumptions (A1), (A3) and (A6) hold for the original data, they also hold for the expanded data given in (13). Because of assumption (C1), we also have assumption (A4) and (A5) for \tilde{X} , thanks to the strict positiveness of the exponential. We can notice that the parameter space is now changed because of the dilatation of the data. The parameter β is the same as the original problem, but the parameter σ is now changed to $\tilde{\sigma}$ with $\tilde{\sigma}_i = \exp\left(\frac{f_{n_i}(\theta)}{n_i}\right) \sigma_i$. We can then define the vector of the true value for parameters in the expanded model (13) as $\tilde{\theta}^0 = (\Phi(\theta^0), \beta^0)$. Using condition (C2), we have $\tilde{\theta}^0$ in the interior of the new parameter set, and so we have assumption

(A2) for the expanded model given in (13).

From Theorem 1, we deduce the asymptotic normality for the maximum likelihood estimator of the expanded data given by (13). We can now consider the optimization problem defining this estimator. The function of $\tilde{\boldsymbol{\theta}}$ to optimize can be written in the following form:

$$\int_{\mathbb{R}^n} \left(\sum_{i=1}^n \sum_{j=1}^{n_i} \rho_{\tau} \left(\frac{\tilde{Y}_{ij} - \tilde{\mu}_{ij}}{\tilde{\sigma}_i} \right) + \sum_{i=1}^n n_i \ln(\tilde{\sigma}_i) \right) d\mathbf{v}$$

with $\tilde{\mu}_{ij} = \tilde{\mathbf{X}}_{ij}^t \boldsymbol{\beta} + \exp\left(\frac{f_{n_i}(\boldsymbol{\theta})}{n_i}\right) e_i^t \mathbf{v}$. We can rewrite this expression as a function of $\boldsymbol{\theta}$:

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \rho_{\tau} \left(\frac{(Y_{ij} - \mu_{ij}) \exp\left(\frac{f_{n_i}(\boldsymbol{\theta})}{n_i}\right)}{\sigma_i \exp\left(\frac{f_{n_i}(\boldsymbol{\theta})}{n_i}\right)} \right) + \sum_{i=1}^n n_i \ln \left(\sigma_i \exp\left(\frac{f_{n_i}(\boldsymbol{\theta})}{n_i}\right) \right)$$

which is exactly equal to (7) when simplified. Therefore, we have the estimator of the expanded problem (13) equal to the estimator $\boldsymbol{\theta}_f$ of $\boldsymbol{\theta}$ by definition. This allows us to conclude of the asymptotic normality of $\boldsymbol{\theta}_f$. Finally, from (20) together with (21) and (22), it follows the form of $\tilde{B}(\boldsymbol{\theta}^0)$. □

We proved the appropriate properties of the estimator, including asymptotic normality, consistency, and the possibility to easily add a term in the optimization problem defining our estimator using expanded data. Our estimation algorithm is carefully described in the following subsection.

2.4 Numerical estimation

Using the equivalence between the maximization of the likelihood and the minimization of the optimization problem defining the quantile estimator, we are therefore able to compute this estimator using classical likelihood optimization algorithms. As mentioned before, we approximate the likelihood given in (5) using Gauss quadrature, via the same lines as in Geraci and Bottai (2014). Their method was coded in an R package called *lqmm* (Geraci, 2014). Based on their

algorithm, we then want to find $\hat{\boldsymbol{\theta}}$ that maximizes the following approximated likelihood defined $l_{app}(\boldsymbol{\theta}|\mathbf{Y})$ as

$$\log \left(\sum_{k_1, \dots, k_n=1}^K \frac{\tau^N (1-\tau)^N}{\prod_{i=1}^n \sigma_i^{n_i}} \exp \left(- \sum_{i=1}^n \sum_{j=1}^{n_i} \rho_\tau \left(\frac{Y_{ij} - \mu_{ij}^{(k_1, \dots, k_n)}}{\sigma_i} \right) \right) f_{\nu}(\mathbf{v}_{k_1, \dots, k_n}) \prod_{l=1}^n w_{k_l} \right) \quad (14)$$

with

$$f_{\nu}(\mathbf{v}_{k_1, \dots, k_n}) = \frac{1}{(2\pi\phi)^{\frac{n}{2}} \sqrt{|\Sigma|}} \exp \left(- \frac{\mathbf{v}_{k_1, \dots, k_n}^t \Sigma^{-1} \mathbf{v}_{k_1, \dots, k_n}}{2\phi} \right),$$

$\mu_{ij}^{(k_1, \dots, k_n)} = \mathbf{X}_{ij}^t \boldsymbol{\beta} - \mathbf{e}_i^t \mathbf{v}_{k_1, \dots, k_n}$, $|\Sigma|$ denoting the determinant of the Σ matrix, $K \in \mathbb{N}^*$ being the order of the quadrature, $\mathbf{v}_{k_1, \dots, k_n} \in \mathbb{R}^n$ being the nodes and $w_{k_i} \in \mathbb{R}$ denoting the weights of the quadrature for all $k_i \in \{1, \dots, k\}$ and $i \in \{1, \dots, n\}$.

As the maximization of (14) does not have an analytical solution, a numerical optimization algorithm is used to solve it. We use either a gradient-search method or a direct search method in which both initial values for $\boldsymbol{\theta}$ and ϕ are required to generate a sequence of improving approximate solutions. For $\boldsymbol{\beta}^{init}$, we use the least squared regression estimate of (1) computed on the data while for σ^{init} we use the link between variance and the scale parameters, as shown in (11), so that we set for all $i \in \{1, \dots, n\}$

$$\sigma_i^{init} = \frac{\tau^2(1-\tau^2)}{(1-2\tau+2\tau^2)} \frac{1}{n_i} \sum_{j=1}^{n_i} \left(Y_{ij} - \mathbf{X}_{ij}^t \boldsymbol{\beta}^{init} - \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \mathbf{X}_{ij}^t \boldsymbol{\beta}^{LS}) \right)^2$$

To initialize ϕ , we use the following estimator:

$$\phi^{init} = \frac{1}{n} \sum_{i=1}^n \left(\left(\Sigma^{-\frac{1}{2}} \mathbf{Q}_\tau(\mathbf{Y} - \mathbf{X}^t \boldsymbol{\beta}^{init}) \right)_i - \frac{1}{n} \sum_{i=1}^n \left(\Sigma^{-\frac{1}{2}} \mathbf{Q}_\tau(\mathbf{Y} - \mathbf{X}^t \boldsymbol{\beta}^{init}) \right)_i \right)^2 \quad (15)$$

with $\mathbf{Q}_\tau(\mathbf{V}) \in \mathbb{R}^n$ being the vector of the τ -th empirical quantiles for all the measures of each individual of the \mathbb{R}^N vector \mathbf{V} .

Once these initial values are computed, the numerical optimization algorithm can be used to infer a first estimate $\hat{\boldsymbol{\theta}}^1$ of $\hat{\boldsymbol{\theta}}$. For the parameter ϕ , we obtain a first a value ϕ^1 by using

equation (15) in which β^{init} is replaced by β^1 . Then, the optimization process is repeated until the various estimates stay unchanged between two consecutive iterations. In order to illustrate the properties of our model, we implement a simulation study based on $N = 1000$ realizations. For more simplicity, we call our models *hlqmm* for heteroscedastic linear quantile mixed model and *phlqmm* for penalized heteroscedastic linear quantile mixed model.

3 Simulation study

This section is devoted to numerical experiments in order to evaluate the performance of our estimates. All our algorithms used in this paper can be found here: <https://github.com/I621974/hlqmm>.

3.1 Asymptotic normality and convergence

We first study the empirical distribution of $\hat{\theta}$ and compare it to the asymptotic theoretical distribution. To compute our estimate, we need to evaluate the asymptotic variance-covariance matrix that involves some conditional expectations with respect to \mathbf{Y} (as shown in Section 2.2). To that end, we use an empirical estimator of conditional expectation with respect to \mathbf{Y} for each $i \leq n$ that can be defined in our case for a generic random variable $\mathbf{Z} \in \mathbb{R}^{n_i}$ as:

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{|S_j|} \sum_{k \in S_j} Z_k$$

where $|S_j|$ denotes the cardinal of S_j defined as the set of indices $\{k | Y_k = Y_{ij}\}$.

To study the empirical distribution of $\hat{\theta}$, we consider the following simulation design in which $p = 9$, $n = 6$, for all $i \in \{1, \dots, n\}$ $n_i = J = 1000$, $\tau = 0.5$ and

$$\psi = 2, \quad \sigma^0 = (1, 2, 1, 3, 1, 0.5), \quad \beta^0 = (1, 2, -2, -1, 3, 5.5, 0.5, 0.1), \quad \phi = 5$$

where ψ denotes the true value for the intercept. Then, we consider three distributions to generate the co-variables \mathbf{X} , namely a normal distribution $\mathcal{N}(2, 2)$, a Laplace distribution $\mathcal{L}(2, 3)$ and a Bernoulli distribution $\mathcal{B}(0.33)$. For each distribution, we represent the frequency histogram for both $\hat{\sigma}$ and $\hat{\beta}$ of 1000 Monte Carlo replications and compare it to the asymptotic theoretical normal distribution. Figure 1 displays the empirical distributions of $\hat{\beta}$ when covariates are generated according to the normal distribution $\mathcal{N}(2, 2)$. All the Figures for the other covariate distributions and parameters can be found in the supplementary material.

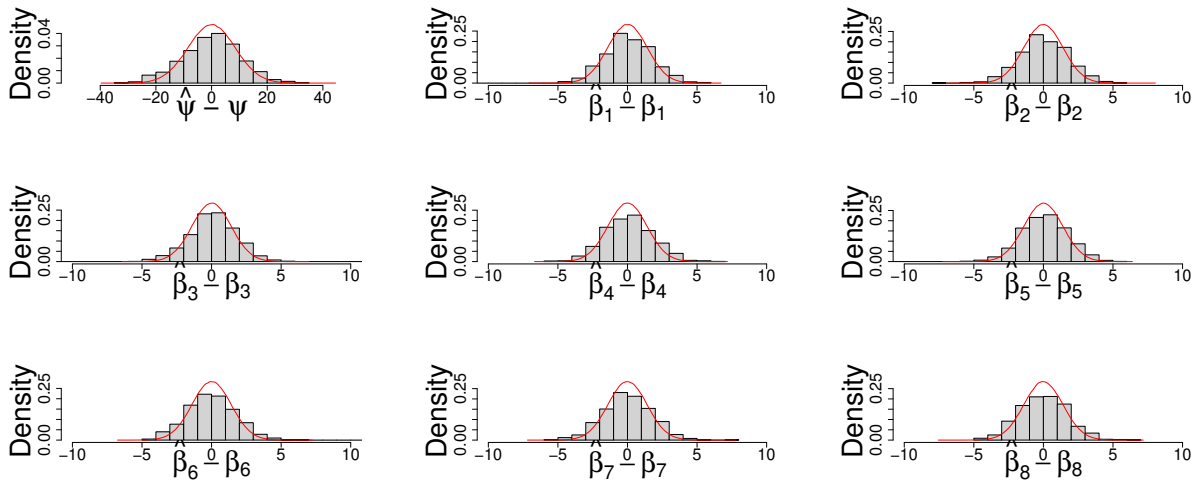


Figure 1: Empirical distribution of $\hat{\beta} - \beta$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$. For the estimator $\hat{\psi}$, the empirical bias was removed.

Firstly, we can see that all empirical distributions fit the theoretical ones, except for the intercept one. There exists a bias on the intercept, which was removed in Figure 1 by centering ϕ . One can find some explanation of this bias, as well as a justification of the use of an intercept in quantile regression in Jurecková and Sen (1996) and Battagliola et al. (2022). Otherwise, we can see that all estimators seem to fit the theoretical normal distribution. In Figure 2, except for some rare extreme values, almost all points indicate a strong match with the asymptotic theoretical distribution for the parameter σ . This was also confirmed by some Shapiro's tests which all

provided p value lower than 5%.

The behavior for other co-variable distributions is mostly the same, except for the Bernoulli

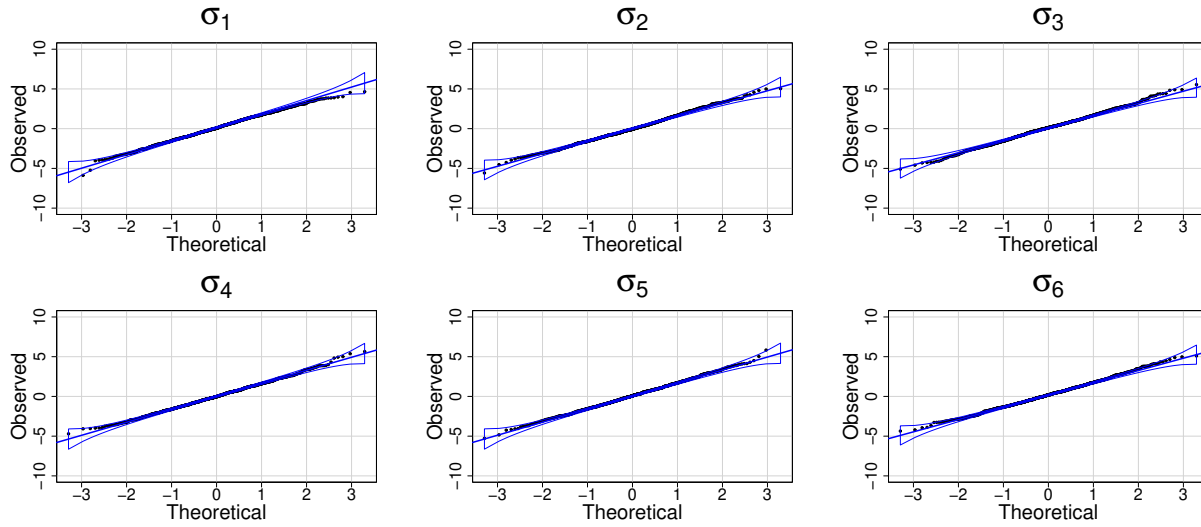


Figure 2: QQ-plot of $\hat{\sigma} - \sigma^0$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$.

generation which presents a higher rate of extreme values. Figures 5 to 19 for these distributions are displayed in the supplementary material.

Next, to study the convergence of our estimator $\hat{\theta}$, we use the same simulation design, but we make the number of measurements, J , vary. We can see the convergence of the algorithm in Figure 3 for parameter σ when J increases. Here, the results are fairly similar, whatever the distribution used to generate the co-variables. Looking at the scale on the ordinate axis, we can see that the closer the parameter is to 0, the harder it is for the method to estimate it correctly. Results are the same for parameter β and for other generations of covariables (as we can see in Figures 5 to 19 in the supplementary material).

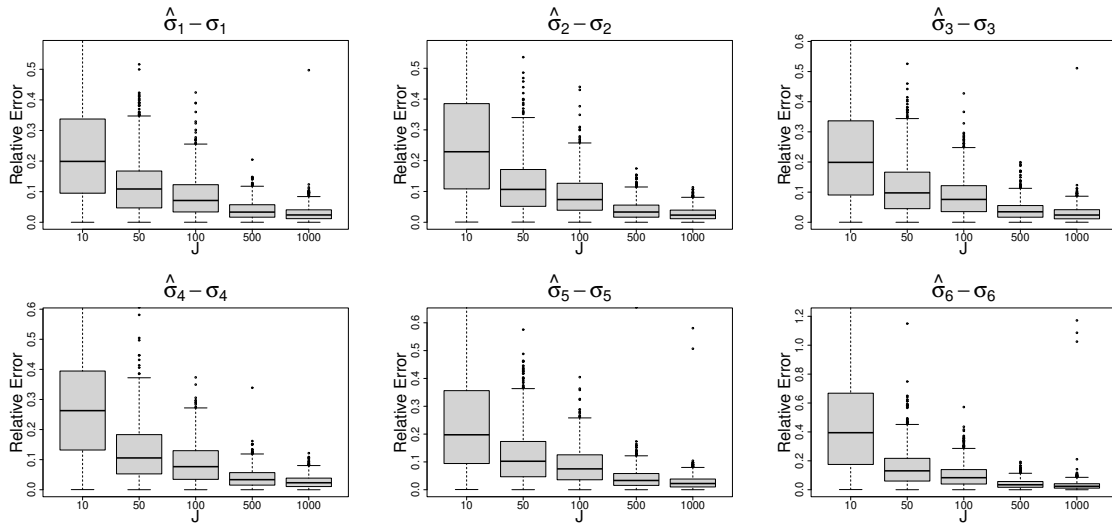


Figure 3: Convergence plot of $\hat{\sigma} - \sigma^0$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$.

4 Application for real data

We consider a data set consisting of isotopic signature measurements on coral reef fish and environmental and biological covariables from the Marquesas Islands (Fey et al. (2021)). Here, we are interested in the isotopic signatures of carbon δ^C and nitrogen δ^N , which provide information on the diet of observed species (Fry (1988), Vander Zanden and Rasmussen (1999)). Precisely, the objective is to predict the trophic diversity of these fish correctly. The trophic diversity, which denotes the diversity of the diet of a species (Newsome et al. (2007)), can be defined by using the size of the isotopic niche. The latter can be calculated from the area of the convex envelope of the measurements when viewed in the (δ^C, δ^N) plane (Fey et al. (2021)). As a high variance can be, at the first order, interpreted as a high diversity, we expect that our method can better predict the isotopic niche's size, with the use or not of the penalized version. Firstly, heteroscedasticity is explicitly included in our models (with and without penalty) through the parameters $\sigma_i, 1 \leq i \leq n$. This gives us access to a measure of diversity for each species. Secondly, the penalized version of our model could compensate for the modeling error, especially when our model without penalty

provides an overestimation of variability. Overall, our models (with and without penalty) provide us with information on trophic diversity, as well as on the links between the isotopic niche and the environmental and biological variables in our dataset.

Regarding the dataset, it contains information about 43 species with enough measurement to compute our model, ranging from 9 to 70 fishes of one species. There are two environmental covariables studied, namely the wind exposition and the season of the measurements, as well as 6 biological covariables, mostly qualitative variables depending only on the species, like a description of the diet or the schooling of the species. As several species are studied in the data set and because we are interested in predicting the trophic diversity of a species, the species are considered as our individuals for which several measurements are available. We apply our method on the dataset and we construct two models: one model to both explain and predict the carbon isotopic signature and a second one to both explain and predict the nitrogen isotropic signature. In these two models, the environmental and biological covariables are used as explanatory variables and Y_{ij} denotes, for the j -th measurement of the i -th species, the model response that is either the carbon isotopic signature or the nitrogen isotropic signature. Finally, we use the predicted values for both the carbon isotopic signature and the nitrogen isotropic signature to predict the trophic diversity of a species.

The isotopic niche for a given species, denoted by T^α , is defined as the convex envelope of the $\alpha\%$ points closest to the barycenter of all the points in the plane (δ^C, δ^N) . It is described in more detail in Fey et al. (2021). Moreover, the area used to compute the size of the isotropic niche for one species heavily depends on the within-species variance, which is the variance of the measurements for the considered species. Therefore, being able to control the within-species variability

of each isotopic signature should enable us to predict more accurately the size of the isotopic niche. In our model (3), the within-species variability is controlled by σ defined in (11). Then, improvement in the computation of the parameter vector σ can help to study the trophic diversity.

Furthermore, our method requires that the variance-covariance matrix Σ , which defines the dependency between each species, to be known. Here, we use the phylogenetic proximities. As in Grafen (1989), the dependency between two species is defined based on a hierarchical classification depending on different phylogenetic scales (family, genus, species). First, we compute a phylogenetic distance h that is evenly distributed on the $[0, 1]$ axis according to whether or not two individuals belong or not to different hierarchical classes (1 if the two species are not from the same family, 0 if the two species are the same). Then, Σ is defined as:

$$(\Sigma)_{i_1, i_2} = 1 - h(i_1, i_2)^\rho \quad i_1, i_2 \in \{1, \dots, n\}$$

where ρ is a fixed parameter, chosen by cross-validation and under the constraint that Σ has to be positive definite. Here, as the dataset contains 43 distinct species, Σ is a 43×43 matrix. This also means that the parameter space Θ is at least of dimension 43, and so makes the estimation problem challenging.

Here, to be able to measure the performance of our methods on the data, we split the data between a training set containing 75% of the data and a testing set. We then perform a Monte-Carlo cross-validation to fix the value of $(\rho, \tau^C, \tau^N, \lambda^C, \lambda^N)$. Determining the order of the τ^C and τ^N quantiles by cross-validation is unusual, but coherent insofar as these parameters are also involved in the (3) equation defining the model, and not just in calculating the estimator $\hat{\theta}$ of equation (6). To validate our model, we simulate the asymptotic distribution of carbon and nitrogen isotope signatures according to our model. We therefore generate data on the same model as the available dataset. By empirically estimating parameters from normal distributions for

quantitative co-variables and Bernoulli distributions for qualitative co-variables, we can simulate isotopic signatures for each species according to equation (3). Random effects can be inferred in the same way as in Geraci and Bottai (2014). We then estimate the theoretical isotopic niche of the species studied for different values of α between 0 and 100. The parameter set chosen will then be that which simulates an isotopic niche of parameter α containing a percentage as close as possible to $\alpha\%$ of the real data from the training set, for all values of α .

Once this is done, we can compare our two methods *hlqmm* and *phlqmm* with the *rqp*d function.

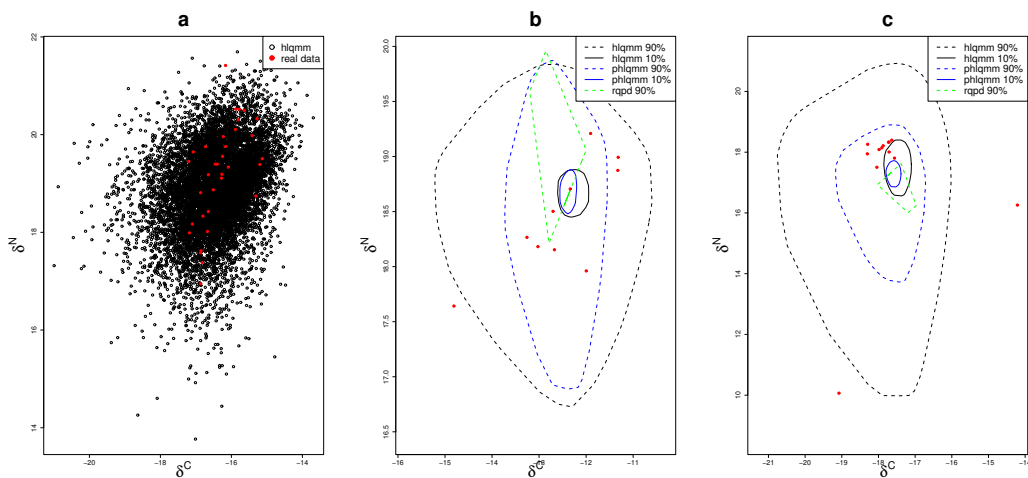


Figure 4: Figure a represents a theoretical isotopic niche simulated with the *hlqmm* model for the species *Epinephelus fasciatus* as well as real data points. Figure b compares the performance of *hlqmm*, *phlqmm* and *rqp*d for the species *Chaetodon ornatissimus*. Figure c compares the performance of *hlqmm*, *phlqmm* and *rqp*d for the species *Thalassoma amblycephalum*.

The test data set is added and the match between the theoretical isotopic niche and the real data is observed again. Figure 4a shows an example of the theoretical isotopic niche, with the simulated points in black and the real data in red. In Figures 4b and 4c, we display the convex envelopes defining the isotopic niche for $\alpha = 90\%$ and $\alpha = 10\%$ for both the three estimators and the real data. We, therefore, want 90% of the real data, represented by red dots, to fall within the largest

of the convex envelopes of a given color, and 10% of this data to fall within the smallest envelope. Figure 4b shows a case where the *hlqmm* method best predicts the data. The penalized method *phlqmm* applies too strong a constraint on the $(\sigma_i)_{1 \leq i \leq n}$, not sufficiently capturing the points furthest from the barycenter. We can see that *hlqmm* performs better as the farthest three red dots on the right are inside the dotted black convex hull but not inside the blue and green one. It's when the *hlqmm* method overestimates within-species variability that the penalized method performs better, as shown in Figure 4c. Here, isotopic niches for small α values are poorly estimated by the *hlqmm* method, and penalization brings the isotopic niche closer to its true shape. We can see that *phlqmm* performs better as the two red dots inside the black convex hull delimited by a solid line are not inside the blue one. In all case, the 10% convex hull of *rqpd* is too little to be seen here. The *rqpd* routine consistently under-performs our two models in terms of isotopic niche matching.

Once we estimate the within-species variability, we can then use σ to study the link between environmental and biological covariables and trophic diversity, with for instance an analysis of variance. The theoretical isotope niche is also interesting in its own right, for example, to identify outliers or to compare the trophic diversity of several different species. As we have seen, our method provides a good description of the dispersion between measurements of the same species, enabling us at the same time to describe the diversity of the diet of coral fish, thus reducing the problem to one of surface estimation. It is also interesting to note that the penalized method fulfills its role well, compensating for any overestimation of dispersion. To improve accuracy, it would also be useful to develop a method to counter the underestimation of certain within-species variability.

5 Discussion

The article developed above presents several modeling approaches for inference in the context of longitudinal data with the presence of heteroscedasticity between individuals. The asymptotic normality properties demonstrated enable a comfortable inference framework to be set up, including the calculation of confidence intervals on the parameters. There are, however, avenues for theoretical improvement that would make this model more complete. For example, the simulations revealed a sensitivity to similarity between within-individual and between-individual variabilities. If the parameter ϕ is too close to elements of the parameter σ , the estimation could be compromised. This weakness of our model therefore calls for accurate and efficient estimation of the ϕ parameter. Unlike the model developed by Geraci and Bottai (2014), whose asymptotic properties were demonstrated in Weidenhammer (2017), we estimate this parameter separately using empirical estimators. It would be interesting to develop a formal framework for estimating this parameter jointly. Since the estimation of this parameter will depend on n and we don't want to make it tend towards infinity (see Section 2.1), a different model should be proposed to make sense of the convergence of the ϕ estimator.

We have given only one example of the class of estimators defined by equation (7) in the form of Ridge penalization of the optimization problem. However, other estimators with different properties could be imagined. We see in Section 4 that penalization has the effect of contracting the surface of isotopic niches, which has the consequence of alleviating problems of overestimation of variability between measurements. In the same way, we could think of a way to overcome the problems of underestimation of dispersion. In general, it would be interesting to study the effect of the choice of the f function in the equation (7) on the value of the θ_f estimator. We could also have defined a lasso penalty in the same way. Although the asymptotic normality

of the θ_f estimator is established in a fairly broad framework, the form of the f function chosen can lead to difficulties in numerical optimization. In the case of Ridge penalization developed in this article, our optimization method can run into local minima quite frequently (around 10% of the time). To solve this problem, more efficient optimization algorithms can be implemented, such as stochastic optimization algorithms (Schneider and Kirkpatrick (2007)).

An application is developed in Section 4 in which the objective is to correctly estimate a surface through the notion of isotopic niches. As mentioned above, the models proposed in this article could equally well be applied to volume estimation problems of any dimension greater than 1, subject to the independence of the variables defining each dimension. This independence assumption may be restrictive in some cases, and it would be interesting to develop a model similar to those presented in this paper but taking into account the dependence between the variables studied. This would require the implementation of multi-dimensional inference, which raises several technical issues. To the best of our knowledge, the definition of asymmetric Laplace's distribution as used in this paper in higher dimensions has not yet been well developed, particularly in a dependency framework. Added to this difficulty is the multiplicity of distributions to which the term multivariate asymmetric Laplace distribution may refer (Kotz et al. (2001)).

In an ecological framework, we have linked the notion of dispersion to that of diversity, but these two notions are very distinct. There are several definitions of diversity which each describe a different reality (Cucherousset and Villéger (2015)) and which it would be interesting to investigate mathematically. The advantage of our model is that it would be possible to control the estimation of diversity according to several of these definitions using the θ_f estimators. One of the conditions to achieve this would be to be able to express the newly defined diversity as a func-

tion of the parameters θ of the model. The notion of diversity is linked to that of specialization in ecology and the latter notion is linked to the adaptation of species to environmental changes (Haaland et al. (2020)). In behavioral ecology, measuring the diversity of behaviors of a species and comparing it to that of other species could be the key to a comparative study of the adaptive capacity of different species sharing an ecosystem.

6 REFERENCES

- Abrevaya, J. and Dahl, C. M. (2008), ‘The effects of birth inputs on birthweight: evidence from quantile estimation on panel data’, *Journal of Business & Economic Statistics* **26**(4), 379–397.
- Aghamohammadi, A. and Mohammadi, S. (2017), ‘Bayesian analysis of penalized quantile regression for longitudinal data’, *Statistical Papers* **58**, 1035–1053.
- Alhamzawi, R. and Yu, K. (2014), ‘Bayesian lasso-mixed quantile regression’, *Journal of Statistical Computation and Simulation* **84**(4), 868–880.
- Arellano, M. and Bonhomme, S. (2013), ‘Random effects quantile regression’, *Retrieved from* .
- Battagliola, M. L., Sørensen, H., Tolver, A. and Staicu, A.-M. (2022), ‘A bias-adjusted estimator in quantile regression for clustered data’, *Econometrics and Statistics* **23**, 165–186.
- Booth, J. G. and Hobert, J. P. (1999), ‘Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **61**(1), 265–285.
- Briollais, L. and Durrieu, G. (2017), ‘Quantile regression for genetic and genomic applications’, *Handbook of Quantile Regression* pp. 409–427.

- Cade, B. S. and Noon, B. R. (2003), 'A gentle introduction to quantile regression for ecologists', *Frontiers in Ecology and the Environment* **1**(8), 412–420.
- Cucherousset, J. and Villéger, S. (2015), 'Quantifying the multiple facets of isotopic diversity: new metrics for stable isotope ecology', *Ecological Indicators* **56**, 152–160.
- Demidenko, E. (1997), Asymptotic properties of nonlinear mixed-effects models, in 'Modelling Longitudinal and Spatially Correlated Data', Springer, pp. 49–62.
- Durrieu, G. and Briollais, L. (2009), 'Sequential design for microarray experiments', *Journal of the American Statistical Association* **104**(486), 650–660.
- Fey, P., Letourneur, Y. and Bonnabel, S. (2021), 'The α -minimum convex polygon as a relevant tool for isotopic niche statistics', *Ecological Indicators* **130**, 108048.
- Friedman, J., Hastie, T. and Tibshirani, R. (2000), 'Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)', *The Annals of Statistics* **28**(2), 337 – 407.
- Fry, B. (1988), 'Food web structure on georges bank from stable c, n, and s isotopic compositions', *Limnology and oceanography* **33**(5), 1182–1190.
- Galarza, C. E., Lachos, V. H. and Bandyopadhyay, D. (2017), 'Quantile regression in linear mixed models: a stochastic approximation em approach', *Statistics and its Interface* **10**(3), 471.
- Galvao, A. F. and Poirier, A. (2019), 'Quantile regression random effects', *Annals of Economics and Statistics* (134), 109–148.
- Geraci, M. (2014), 'Linear quantile mixed models: the lqmm package for laplace quantile regression', *Journal of Statistical Software* **57**, 1–29.

- Geraci, M. and Bottai, M. (2007), ‘Quantile regression for longitudinal data using the asymmetric laplace distribution’, *Biostatistics* **8**(1), 140–154.
- Geraci, M. and Bottai, M. (2014), ‘Linear quantile mixed models’, *Statistics and computing* **24**, 461–479.
- Grafen, A. (1989), ‘The phylogenetic regression’, *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **326**(1233), 119–157.
- Haaland, T. R., Wright, J. and Ratikainen, I. I. (2020), ‘Generalists versus specialists in fluctuating environments: A bet-hedging perspective’, *Oikos* **129**(6), 879–890.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (2011), *Robust statistics: the approach based on influence functions*, John Wiley & Sons.
- Hubert, M. and Branden, K. V. (2003), ‘Robust methods for partial least squares regression’, *Journal of Chemometrics: A Journal of the Chemometrics Society* **17**(10), 537–549.
- Ji, Y. and Shi, H. (2022), ‘Shrinkage estimation of fixed and random effects in linear quantile mixed models’, *Journal of Applied Statistics* **49**(14), 3693–3716.
- Jiang, J. (1998), ‘Asymptotic properties of the empirical blup and blue in mixed linear models’, *Statistica Sinica* pp. 861–885.
- Jiang, J. (2017), *Asymptotic analysis of mixed effects models: theory, applications, and open problems*, CRC press.
- Jurecková, J. and Sen, P. K. (1996), *Robust statistical procedures: asymptotics and interrelations*, Vol. 311, John Wiley & Sons.
- Kim, M.-O. and Yang, Y. (2011), ‘Semiparametric approach to a random effects quantile regression model’, *Journal of the American Statistical Association* **106**(496), 1405–1417.

- Koenker, R. (1994), Confidence intervals for regression quantiles, in 'Asymptotic statistics: proceedings of the fifth prague symposium, held from September 4–9, 1993', Springer, pp. 349–359.
- Koenker, R. (2004), 'Quantile regression for longitudinal data', *Journal of multivariate analysis* **91**(1), 74–89.
- Koenker, R. (2005), *Quantile regression*, Vol. 38, Cambridge university press.
- Koenker, R. (2017), 'Quantile regression: 40 years on', *Annual review of economics* **9**, 155–176.
- Koenker, R. and Bassett, J. G. (1978), 'Regression quantiles', *Econometrica: journal of the Econometric Society* pp. 33–50.
- Koenker, R., Chernozhukov, V., He, X. and Peng, L. (2017), 'Handbook of quantile regression'.
- Kotz, S., Kozubowski, T. J., Podgórski, K., Kotz, S., Kozubowski, T. J. and Podgórski, K. (2001), *Asymmetric multivariate Laplace distribution*, Springer.
- Lamarche, C. (2010), 'Robust penalized quantile regression estimation for panel data', *Journal of Econometrics* **157**(2), 396–408.
- Liang, K.-Y. and Zeger, S. L. (1986), 'Longitudinal data analysis using generalized linear models', *Biometrika* **73**(1), 13–22.
- Lu, X. and Fan, Z. (2015), 'Weighted quantile regression for longitudinal data', *Computational Statistics* **30**, 569–592.
- Luo, Y., Lian, H. and Tian, M. (2012), 'Bayesian quantile regression for longitudinal data models', *Journal of Statistical Computation and Simulation* **82**(11), 1635–1649.

- Miller, J. J. (1977), 'Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance', *The Annals of Statistics* pp. 746–762.
- Newsome, S. D., Martinez del Rio, C., Bearhop, S. and Phillips, D. L. (2007), 'A niche for isotopic ecology', *Frontiers in Ecology and the Environment* **5**(8), 429–436.
- Peng, L., Wu, J. and Benatar, M. (2009), 'Developing reference data for nerve conduction studies: an application of quantile regression', *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine* **40**(5), 763–771.
- Pinheiro, J. C. (1994), *Topics in mixed effects models*, The University of Wisconsin-Madison.
- Schneider, J. and Kirkpatrick, S. (2007), *Stochastic optimization*, Springer Science & Business Media.
- Sheather, S. J. and Marron, J. S. (1990), 'Kernel quantile estimators', *Journal of the American Statistical Association* **85**(410), 410–416.
- Steyer, R. (2013), *Wahrscheinlichkeit und regression*, Springer-Verlag.
- Stigler, S. M. (1973), 'Simon newcomb, percy daniell, and the history of robust estimation 1885–1920', *Journal of the American Statistical Association* **68**(344), 872–879.
- Tian, Y., Wang, L., Tang, M., Zang, Y. and Tian, M. (2020), 'Likelihood-based quantile autoregressive distributed lag models and its applications', *Journal of Applied Statistics* **47**(1), 117–131.
- Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.
- Vander Zanden, M. J. and Rasmussen, J. B. (1999), 'Primary consumer $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ and the trophic position of aquatic consumers', *Ecology* pp. 1395–1404.

- Wang, S. and Xiang, L. (2017), 'Two-layer em algorithm for ald mixture regression models: A new solution to composite quantile regression', *Computational Statistics & Data Analysis* **115**, 136–154.
- Weidenhammer, B. (2017), *The Consistency of Quantile Regression in Linear Mixed Models*, Freie Universitaet Berlin (Germany).
- Weiss, L. (1971), 'Asymptotic properties of maximum likelihood estimators in some nonstandard cases', *Journal of the American Statistical Association* pp. 345–350.
- Weiss, L. (1973), 'Asymptotic properties of maximum likelihood estimators in some nonstandard cases, ii', *Journal of the American Statistical Association* pp. 428–430.
- Wu, Q. and Yao, W. (2016), 'Mixtures of quantile regressions', *Computational Statistics & Data Analysis* **93**, 162–176.
- Yu, K., Lu, Z. and Stander, J. (2003), 'Quantile regression: applications and current research areas', *Journal of the Royal Statistical Society Series D: The Statistician* **52**(3), 331–350.
- Yu, K. and Moyeed, R. A. (2001), 'Bayesian quantile regression', *Statistics & Probability Letters* **54**(4), 437–447.
- Yu, K. and Zhang, J. (2005), 'A three-parameter asymmetric laplace distribution and its extension', *Communications in Statistics—Theory and Methods* **34**(9-10), 1867–1879.

SUPPLEMENTARY MATERIAL

Proofs

Computation of the second derivatives of the log-likelihood

It follows from assumption (A3)

$$l(\eta|\mathbf{Y}) = \log \left(\int_{\mathbb{R}^n} f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv \right)$$

with

$$f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) = \prod_{i=1}^n \prod_{j=1}^{n_i} \frac{\tau(1-\tau)}{\sigma_i} \exp \left(-\rho_{\tau} \left(\frac{Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v}{\sigma_i} \right) \right)$$

and

$$f_{\nu}(v) = \frac{1}{(2\pi\phi)^{\frac{n}{2}} \sqrt{|\Sigma|}} \exp \left(-\frac{v^t \Sigma^{-1} v}{2\phi} \right).$$

Consequently, we have for $\boldsymbol{\theta} \in \{\boldsymbol{\beta}, \boldsymbol{\sigma}\}$

$$\frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\theta}} = \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \int_{\mathbb{R}^n} \frac{\partial f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)}{\partial \boldsymbol{\theta}} f_{\nu}(v) dv. \quad (16)$$

First derivatives

We have :

$$\begin{aligned} \frac{\partial f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)}{\partial \boldsymbol{\sigma}} &= \tau^N (1-\tau)^N \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\sigma}} \left(\frac{1}{\sigma_i} \exp \left(-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) \right) \\ &\times \prod_{k=1, k \neq i}^n \frac{1}{\sigma_k} \exp \left(-\frac{1}{\sigma_k} \sum_{j=1}^{n_k} \rho_{\tau} (Y_{kj} - X_{kj}^t \boldsymbol{\beta} - z_k^t v) \right). \end{aligned}$$

Moreover,

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{\sigma}} \left(\frac{1}{\sigma_i} \exp \left(-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) \right) = \\ &\frac{\partial}{\partial \boldsymbol{\sigma}} \left(\frac{1}{\sigma_i} \right) \exp \left(-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) \\ &+ \frac{1}{\sigma_i} \frac{\partial}{\partial \boldsymbol{\sigma}} \left(\exp \left(-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) \right). \end{aligned}$$

Consequently, it follows that

$$\frac{\partial f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)}{\partial \boldsymbol{\sigma}} = \sum_{i=1}^n \frac{\mathbf{e}_i}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v),$$

with \mathbf{e}_i the i -th unit vector of \mathbb{R}^n .

In the end, we have:

$$\frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\sigma}} = \mathbb{E} \left[\sum_{i=1}^n \frac{\mathbf{e}_i}{\sigma_i^2} \sum_{j=1}^{n_i} \rho_{\tau}(\epsilon_{ij}) \middle| \mathbf{Y} \right] - \sum_{i=1}^n \frac{\mathbf{e}_i}{\sigma_i}$$

by definition of the linear model and Bayes formula.

We have :

$$\begin{aligned} \frac{\partial f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)}{\partial \boldsymbol{\beta}} &= \tau^N (1-\tau)^N \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} \left(\frac{1}{\sigma_i} \exp \left(-\frac{1}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) \right) \\ &\times \prod_{k=1, k \neq i}^n \frac{1}{\sigma_k} \exp \left(-\frac{1}{\sigma_k} \sum_{j=1}^{n_k} \rho_{\tau}(Y_{kj} - X_{kj}^t \boldsymbol{\beta} - z_k^t v) \right). \end{aligned}$$

Moreover,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} (\rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v)) &= \frac{\partial}{\partial \boldsymbol{\beta}} (\tau - \mathbb{1}_{Y_{ij} \leq X_{ij}^t \boldsymbol{\beta} + z_i^t v}) (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \\ &+ (\tau - \mathbb{1}_{Y_{ij} \leq X_{ij}^t \boldsymbol{\beta} + z_i^t v}) \frac{\partial}{\partial \boldsymbol{\beta}} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \\ &= (\mathbb{1}_{Y_{ij} \leq X_{ij}^t \boldsymbol{\beta} + z_i^t v} - \tau) X_{ij}. \end{aligned}$$

Therefore, we have:

$$\frac{\partial f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{Y_{ij} \leq X_{ij}^t \boldsymbol{\beta} + z_i^t v}) X_{ij} f_{\mathbf{Y}|\nu}(\mathbf{Y}|v).$$

Finally :

$$\frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\beta}} = \tau \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} X_{ij} - \mathbb{E} \left[\sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} \mathbb{1}_{\epsilon_{ij} \leq 0} X_{ij} \middle| \mathbf{Y} \right].$$

For the rest of the calculations, it is useful to determine the various derivatives $\frac{\partial f_{\mathbf{Y}}(\mathbf{Y})}{\partial \boldsymbol{\theta}}$ for $\boldsymbol{\theta} \in$

$\{\boldsymbol{\beta}, \boldsymbol{\sigma}\}$. Given the previous calculations, we have :

$$\frac{\partial f_{\mathbf{Y}}(\mathbf{Y})}{\partial \boldsymbol{\sigma}} = \int_{\mathbb{R}^n} \sum_{i=1}^n \frac{\mathbf{e}_i}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(\epsilon_{ij}) \right) - \sigma_i \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv,$$

and

$$\frac{\partial f_{\mathbf{Y}}(\mathbf{Y})}{\partial \boldsymbol{\beta}} = \int_{\mathbb{R}^n} \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ij} f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv.$$

Derivative with respect to (σ, σ) .

For $i \in \{1, \dots, n\}$, the components of the vector $\frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\sigma}}$ are:

$$\int_{\mathbb{R}^n} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) f_{\nu|\mathbf{Y}}(v|\mathbf{Y}) dv.$$

If we derive this vector by $\boldsymbol{\sigma}$, we create a matrix whose elements are for $i \in \{1, \dots, n\}$ and $k \in \{1, \dots, n\}$:

$$\frac{\partial}{\partial \sigma_k} \left(\int_{\mathbb{R}^n} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) f_{\nu|\mathbf{Y}}(v|\mathbf{Y}) dv \right).$$

For all $i, k \in \{1, \dots, n\}$, this derivative can be written as:

$$\begin{aligned} & \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \frac{\partial}{\partial \sigma_k} (f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)) f_{\nu}(v) dv \\ & + \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \frac{\partial}{\partial \sigma_k} \left(\frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv \\ & - \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}^2(\mathbf{Y})} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) \frac{\partial}{\partial \sigma_k} (f_{\mathbf{Y}}(\mathbf{Y})) dv. \end{aligned} \tag{17}$$

We have:

$$\begin{aligned} & \frac{\partial}{\partial \sigma_k} \left(\frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau} (Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \right) \\ & = \frac{1}{\sigma_i^2} \left(\frac{2}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (\epsilon_{ij}) + 1 \right) \mathbb{1}_{k=i}. \end{aligned}$$

The second term of equation (17) becomes:

$$\mathbb{E} \left[\frac{1}{\sigma_i^2} \left(\frac{2}{\sigma_i} \sum_{j=1}^{n_i} \rho_{\tau} (\epsilon_{ij}) + 1 \right) \mathbb{1}_{k=i} \middle| \mathbf{Y} \right].$$

Subtracting the first and last terms of equation (17), we obtain:

$$\text{Cov} \left(\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \frac{1}{\sigma_k^2} \sum_{l=1}^{n_k} \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) \middle| \mathbf{Y} \right).$$

Finally, we have the following form:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\sigma}} \frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\sigma}} &= \sum_{i=1}^n \frac{\mathbf{e}_i \mathbf{e}_i^t}{\sigma_i^2} + \mathbb{E} \left[\sum_{i=1}^n \frac{2\mathbf{e}_i \mathbf{e}_i^t}{\sigma_i^3} \sum_{j=1}^{n_i} \epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) \middle| \mathbf{Y} \right] \\ &+ \sum_{i=1}^n \sum_{k=1}^n \frac{\mathbf{e}_i \mathbf{e}_i^t}{\sigma_i^2 \sigma_k^2} \text{Cov} \left(\sum_{j=1}^{n_i} \epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \sum_{l=1}^{n_k} \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) \middle| \mathbf{Y} \right). \end{aligned}$$

Derivative with respect to (σ, β) .

If we derive the vector $\frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\sigma}}$ by $\boldsymbol{\beta}$, we create a matrix whose elements are for all $i \in \{1, \dots, n\}$

and for all $k \in \{1, \dots, p\}$

$$\begin{aligned} &\int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \frac{\partial}{\partial \beta_k} (f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)) f_{\nu}(v) dv \\ &+ \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \frac{\partial}{\partial \beta_k} \left(\frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv \\ &- \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}^2(\mathbf{Y})} \frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) \frac{\partial}{\partial \beta_k} (f_{\mathbf{Y}}(\mathbf{Y})) dv. \end{aligned} \tag{18}$$

We have:

$$\begin{aligned} &\frac{\partial}{\partial \beta_k} \left(\frac{1}{\sigma_i^2} \left(\left(\sum_{j=1}^{n_i} \rho_{\tau}(Y_{ij} - X_{ij}^t \boldsymbol{\beta} - z_i^t v) \right) - \sigma_i \right) \right) \\ &= \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk}. \end{aligned}$$

The second term of equation (18) therefore becomes:

$$\mathbb{E} \left[\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} \middle| \mathbf{Y} \right].$$

Subtracting the first and last terms, we obtain:

$$\text{Cov} \left(\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \sum_{p=1}^n \frac{1}{\sigma_p} \sum_{q=1}^{n_p} \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} \middle| \mathbf{Y} \right).$$

Finally, we obtain:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} \frac{\partial l(\eta|\mathbf{Y})}{\partial \boldsymbol{\sigma}} &= \mathbb{E} \left[\sum_{i=1}^n \sum_{k=1}^p \frac{\mathbf{e}_i \mathbf{e}_i^t}{\sigma_i^2} \sum_{j=1}^{n_i} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} \middle| \mathbf{Y} \right] \\ &+ \sum_{i=1}^n \sum_{k=1}^p \mathbf{e}_i \mathbf{e}_i^t \text{Cov} \left(\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \sum_{p=1}^n \frac{1}{\sigma_p} \sum_{q=1}^{n_p} \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} \middle| \mathbf{Y} \right). \end{aligned}$$

Derivative with respect to (β, β) .

If we derive the vector $\frac{\partial l(\eta|\mathbf{Y})}{\partial \beta}$ by β , we create a matrix whose elements are for all $i \in \{1, \dots, n\}$,

and for all $k \in \{1, \dots, p\}$:

$$\begin{aligned} & \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ijk} \frac{\partial}{\partial \beta_l} (f_{\mathbf{Y}|\nu}(\mathbf{Y}|v)) f_{\nu}(v) dv \\ & + \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}(\mathbf{Y})} \frac{\partial}{\partial \beta_l} \left(\sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ijk} \right) f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) dv \\ & - \int_{\mathbb{R}^n} \frac{1}{f_{\mathbf{Y}}^2(\mathbf{Y})} \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ijk} f_{\mathbf{Y}|\nu}(\mathbf{Y}|v) f_{\nu}(v) \frac{\partial}{\partial \beta_l} (f_{\mathbf{Y}}(\mathbf{Y})) dv. \end{aligned} \quad (19)$$

We have :

$$\frac{\partial}{\partial \beta_l} \left(\sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ijk} \right) = 0$$

so the second term of (19) is zero.

Subtracting the first and last terms, we have:

$$\text{Var} \left(\sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) X_{ij} \middle| \mathbf{Y} \right).$$

Finally, simplifying the constant terms, we have :

$$\frac{\partial}{\partial \beta} \frac{\partial l(\eta|\mathbf{Y})}{\partial \beta} = \text{Var} \left(\sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j=1}^{n_i} \mathbb{1}_{\epsilon_{ij} \leq 0} X_{ij} \middle| \mathbf{Y} \right).$$

Determination of the asymptotic variance-covariance matrix

Determination of $B_{\sigma, \sigma}(\theta^0)$.

The second derivative of the likelihood with respect to the parameter σ is a matrix of size $n \times n$

whose element (i, k) is:

$$\begin{aligned} & \left(\frac{1}{\sigma_i^2} + \frac{2}{\sigma_i^3} \sum_{j=1}^{n_i} \mathbb{E} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) \middle| \mathbf{Y}] \right) \mathbb{1}_{i=k} \\ & + \frac{1}{\sigma_i^2 \sigma_k^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) \middle| \mathbf{Y}). \end{aligned}$$

We temporarily note $W_j^i = \mathbb{E} [\epsilon_{ij}(\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}]$ for a fixed $i \in \{1, \dots, n\}$. We have

$$\mathbb{E} [W_j^i] = \tau \mathbb{E} [\epsilon_{ij}] - \mathbb{E} [\epsilon_{ij} \mathbb{1}_{\epsilon_{ij} \leq 0}] = \frac{\sigma_i(1 - 2\tau)}{(1 - \tau)} + \frac{\tau \sigma_i}{1 - \tau} = \sigma_i.$$

Moreover,

$$\text{Var} (W_j^i) = \mathbb{E} [(W_j^i)^2] - \sigma_i^2 \leq 4\mathbb{E} [\epsilon_{ij}^2(\tau - \mathbb{1}_{\epsilon_{ij} \leq 0})^2] < +\infty$$

because W_j^i is a conditional expectation of a random variable with moments of order two. From the law of large numbers, we have, for all i ,

$$\frac{1}{N} \sum_{j=1}^{n_i} W_j^i = \frac{n_i}{N} \frac{1}{n_i} \sum_{j=1}^{n_i} W_j^i \xrightarrow{P} \sigma_i$$

using assumption **(A1)**.

We note this time for all i, k , $W_{jl}^{ik} = \text{Cov} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})$. We have by independence of the ϵ_{ij} :

$$\begin{aligned} \mathbb{E} [W_{jl}^{ik}] &= \text{Cov} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0})) \\ &\quad - \text{Cov} (\mathbb{E} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}], \mathbb{E} [\epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}]) \\ &= (\text{Var} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0})) - \text{Var} (\mathbb{E} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}])) \mathbb{1}_{k=i} \mathbb{1}_{j=l}. \end{aligned}$$

But we have:

$$\text{Var} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0})) = \sigma_i^2$$

and

$$\text{Var} (\mathbb{E} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}]) = \mathbb{E} [\mathbb{E}^2 [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}]] - \sigma_i^2.$$

At the end, we have :

$$\mathbb{E} [W_{jl}^{ik}] = (2\sigma_i^2 - \mathbb{E} [\mathbb{E}^2 [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]]) \mathbb{1}_{k=i} \mathbb{1}_{j=l}$$

because of the identic distribution of $(\epsilon_{ij})_j$ for a fixed i .

Using similar formulas, and using the fact that the asymmetric Laplace distribution admits moments of infinite order, we can see that $\text{Var}(W_{jl}) \leq M < +\infty$. This is well within the scope of

the weak law of large numbers and we have:

$B_{\sigma,\sigma}(\theta^0) = (b_{\sigma,\sigma}^{i,k})_{i,k}$ where :

$$b_{\sigma,\sigma}^{i,k} = \frac{\mathbb{E}[\mathbb{E}^2[\epsilon_{i1}(\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]]|_{\theta=\theta^0} \mathbb{1}_{k=i}}{(\sigma_i^0)^4} \quad (20)$$

Determination of $B_{\sigma,\beta}(\theta^0)$

The matrix $\frac{\partial}{\partial \beta} \frac{\partial l(\eta|\mathbf{Y})}{\partial \sigma}$ of dimension $n \times p$ have its elements $(i, k), i \in \{1, \dots, n\}, k \in \{1, \dots, p\}$

written:

$$\begin{aligned} &= \mathbb{E} \left[\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} \middle| \mathbf{Y} \right] \\ &+ \text{Cov} \left(\frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} \epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \sum_{p=1}^n \frac{1}{\sigma_p} \sum_{q=1}^{n_p} \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} \middle| \mathbf{Y} \right). \end{aligned}$$

First, we set $W_j^i = \mathbb{E}[(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}]$. We have :

$$\mathbb{E}[W_j^i] = (\mathbb{E}[\mathbb{1}_{\epsilon_{ij} \leq 0}] - \tau) X_{ijk} = 0$$

as $\mathbb{E}[\mathbb{1}_{\epsilon_{ij} \leq 0}] = \tau$. Then, we have:

$$\frac{1}{N} \sum_{j=1}^{n_i} \mathbb{E}[(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) | \mathbf{Y}] X_{ijk} \longrightarrow 0$$

in probability when $N \rightarrow +\infty$.

The second term can be rewritten as follows:

$$\frac{1}{\sigma_i^2} \sum_{p=1}^n \frac{1}{\sigma_p} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \text{Cov}(\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y}).$$

This time, let W_{jq}^{ik} be $\text{Cov}(\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y})$. We have :

$$\begin{aligned} \mathbb{E}[W_{jq}^{ik}] &= \text{Cov}(\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk}) \\ &- \text{Cov}(\mathbb{E}[\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) | \mathbf{Y}], \mathbb{E}[\mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y}]) \\ &= 0 + \mathbb{E}[\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau)] \mathbb{E}[\mathbb{1}_{\epsilon_{ij} \leq 0}] X_{ijk} \\ &- \mathbb{E}[\mathbb{E}[\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) | \mathbf{Y}] \mathbb{E}[\mathbb{1}_{\epsilon_{ij} \leq 0} X_{ijk} | \mathbf{Y}]] \mathbb{1}_{p=i} \mathbb{1}_{j=q} \\ &= (-\tau \sigma_i - \mathbb{E}[\mathbb{E}[\epsilon_{i1} (\mathbb{1}_{\epsilon_{i1} \leq 0} - \tau) | \mathbf{Y}] \mathbb{E}[\mathbb{1}_{\epsilon_{i1} \leq 0} | \mathbf{Y}]]]) X_{ijk} \mathbb{1}_{p=i} \mathbb{1}_{j=q}. \end{aligned}$$

Using assumption (A4), we have convergence of the series $\left(\sum_{j,q} \frac{\mathbb{E}[W_{jq}^{ik}]}{N}\right)$. As before, the variance is finite because we only deal with L^2 random variables. We then have the result:

$B_{\sigma,\beta}(\theta^0) = (b_{\sigma,\beta}^{i,k})_{i,k}$ of size $n \times p$ where :

$$b_{\sigma,\beta}^{i,k} = \frac{c_{ik}}{(\sigma_i^0)^2} \left(\tau + \frac{1}{\sigma_i^0} \mathbb{E} [\mathbb{E} [\epsilon_{i1} (\mathbb{1}_{\epsilon_{i1} \leq 0} - \tau) | \mathbf{Y}] \mathbb{E} [\mathbb{1}_{\epsilon_{i1} \leq 0} | \mathbf{Y}]] |_{\theta=\theta^0} \right) \quad (21)$$

with $c_i = (c_{i1}, \dots, c_{ip}) = \sum_j^{+\infty} \frac{X_{ij}}{N}$.

Determination of $B_{\beta,\beta}(\theta^0)$

We have:

$$\frac{\partial}{\partial \beta} \frac{\partial l(\eta | \mathbf{Y})}{\partial \beta} = \sum_{i=1}^n \sum_{k=1}^n \frac{1}{\sigma_i \sigma_k} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} X_{ij} \text{Cov}(\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} | \mathbf{Y}) X_{kl}^t.$$

Using $W_{jl}^{ik} = X_{ij} \text{Cov}(\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} | \mathbf{Y}) X_{kl}^t$, we have :

$$\begin{aligned} \mathbb{E} [W_{jl}^{ik}] &= X_{ij} (\text{Cov}(\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0}) - \text{Cov}(\mathbb{E}[\mathbb{1}_{\epsilon_{ij} \leq 0} | \mathbf{Y}], \mathbb{E}[\mathbb{1}_{\epsilon_{kl} \leq 0} | \mathbf{Y}])) X_{kl}^t \\ &= X_{ij} (\text{Var}(\mathbb{1}_{\epsilon_{ij} \leq 0}) - \mathbb{E}[\mathbb{E}^2[\mathbb{1}_{\epsilon_{ij} \leq 0} | \mathbf{Y}]] + \tau^2)) X_{ij}^t \mathbb{1}_{k=i} \mathbb{1}_{j=l} \\ &= X_{ij} (\tau - \mathbb{E}[\mathbb{E}^2[\mathbb{1}_{\epsilon_{ij} \leq 0} | \mathbf{Y}]])) X_{ij}^t \mathbb{1}_{k=i} \mathbb{1}_{j=l} \end{aligned}$$

Since this expectation only depends on j or l at the level of the data X_i or X_k , we have convergence of the $\left(\sum_{i,j} \frac{\mathbb{E}[W_{ij}^{ik}]}{N}\right)$ series, using (A5) assumption.

Moreover, the variance of W_{jl}^{ik} is bounded for the same reasons as above. The result is as follows:

$$B_{\beta,\beta}(\theta^0) = \sum_{i=1}^n \frac{\mathbb{E}[\mathbb{E}^2[\mathbb{1}_{\epsilon_{ij} \leq 0} | \mathbf{Y}]] - \tau}{(\sigma_i^0)^2} C_i \quad (22)$$

with C_i a positive definite matrix of size $p \times p$.

Determination of Φ_k for $k \in \{1, \dots, 4\}$

Due to the similarity with the first case, the calculation for (σ, σ) , the other parts of the demonstration is dealt with more briefly.

Determination for (σ, σ)

Φ_1 :

Φ_1 is a matrix of size $n \times n$ whose elements i, k are:

$$\begin{aligned} & \sup_{\theta^1 \in N_n(\theta^0)} \left\{ \frac{1}{N^2} \left(\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^1)^2} \right) + \frac{2\mathbb{1}_{i=k}}{(\sigma_i^2)^3} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}] \right. \\ & - \frac{2\mathbb{1}_{i=k}}{(\sigma_i^1)^3} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^1} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}] \\ & + \frac{1}{(\sigma_i^2 \sigma_k^2)^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov}_{\theta^2} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}) \\ & \left. - \frac{1}{(\sigma_i^1 \sigma_k^1)^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov}_{\theta^1} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}) \right\}. \end{aligned}$$

We have :

$$\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^1)^2} = \frac{(\sigma_i^1 - \sigma_i^2)(\sigma_i^1 + \sigma_i^2)}{(\sigma_i^1)^2(\sigma_i^2)^2}.$$

As $\theta_1, \theta_2 \in N_N(\theta^0)$, we have :

$$|\sigma_i^1 - \sigma_i^0| \leq \frac{M}{N} \quad \text{et} \quad |\sigma_i^2 - \sigma_i^0| \leq \frac{M}{N}$$

and so:

$$\sup_{\theta^1 \in N_n(\theta^0)} |\sigma_i^2 - \sigma_i^1| \leq 2 \frac{M}{N}.$$

We then have:

$$\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^1)^2} \leq 2 \frac{M}{N} C$$

with C an unknown constant.

We saw before that:

$$\frac{1}{N^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}] = o\left(\frac{1}{N}\right)$$

for any θ and:

$$\begin{aligned} & \frac{1}{N^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov}_\theta (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}) \\ & \xrightarrow{N \rightarrow +\infty} \sigma_i^2 \mathbb{1}_{k=i} + \sigma_i \sigma_k - \mathbb{E} [\mathbb{E} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \mathbb{E} [\epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}]] \end{aligned}$$

The last difference defining Φ_1 then tends to:

$$\begin{aligned} & \left(\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^1)^2} \right) \mathbb{1}_{k=i} + \frac{1}{\sigma_i^2 \sigma_k^2} - \frac{1}{\sigma_i^1 \sigma_k^1} \\ & + \left(\frac{1}{(\sigma_i^1)^2 (\sigma_k^1)^2} - \frac{1}{(\sigma_i^2)^2 (\sigma_k^2)^2} \right) \mathbb{E}_{\theta_0} [\mathbb{E}_{\theta_0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \mathbb{E}_{\theta_0} [\epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}]] \end{aligned}$$

because as $\theta_1, \theta_2 \in N_N(\theta_0)$, we have $\mathbb{E}_{\theta_1} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \xrightarrow{N \rightarrow +\infty} \mathbb{E}_{\theta_0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]$

and the same goes for the others expectations.

We have:

$$\frac{1}{\sigma_i^2 \sigma_k^2} - \frac{1}{\sigma_i^1 \sigma_k^1} = \frac{\sigma_i^1 \sigma_k^1 - \sigma_i^2 \sigma_k^2}{\sigma_i^2 \sigma_k^2 \sigma_i^1 \sigma_k^1} = \frac{\sigma_i^1 (\sigma_k^1 - \sigma_k^2) + (\sigma_i^1 - \sigma_i^2) \sigma_k^2}{\sigma_i^2 \sigma_k^2 \sigma_i^1 \sigma_k^1} = \frac{\sigma_k^1 - \sigma_k^2}{\sigma_i^2 \sigma_k^2 \sigma_k^1} + \frac{\sigma_i^1 - \sigma_i^2}{\sigma_i^2 \sigma_i^1 \sigma_k^1}$$

and so:

$$\frac{1}{\sigma_i^2 \sigma_k^2} - \frac{1}{\sigma_i^1 \sigma_k^1} \leq \frac{M}{N} C.$$

In the same way, we have:

$$\left(\frac{1}{(\sigma_i^1)^2 (\sigma_k^1)^2} - \frac{1}{(\sigma_i^2)^2 (\sigma_k^2)^2} \right) \leq \frac{M}{N} C.$$

All these inequalities being true for all $\theta^1 \in N_N(\theta^0)$, we can move on to the sup. Consequently,

we have $M^2 \Phi_1 = o(\frac{M^3}{N^3})$ which tends to 0 by construction of M .

Φ_2 :

We have Φ_2 which is a matrix of size $n \times n$ whose elements i, k are:

$$\begin{aligned} & \frac{1}{N^2} \left(\frac{2\mathbb{1}_{i=k}}{(\sigma_i^2)^3} \left(\sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}] - \mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})] \right) \right. \\ & + \frac{1}{(\sigma_i^2 \sigma_k^2)^2} \left(\sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov}_{\theta^2} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}) \right. \\ & \left. \left. - n_i n_k \left((\sigma_i^2)^2 \mathbb{1}_{k=i} + \sigma_i^2 \sigma_k^2 - \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \mathbb{E}_{\theta^2} [\epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}]] \right) \right) \right). \end{aligned}$$

We set $X_N^{ik} = \frac{1}{N^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Cov}_{\theta^2} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})$.

We have :

$$\mathbb{E}_{\theta^2} [X_N^{ik}] = \frac{n_i n_k}{N^2} ((\sigma_i^2)^2 \mathbb{1}_{k=i} + \sigma_i^2 \sigma_k^2 - \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \mathbb{E}_{\theta^2} [\epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}]]).$$

The last two terms can therefore be written as a constant of N multiplied by the following value:

$$X_N^{ik} - \mathbb{E}_{\theta^2} [X_N^{ik}].$$

By Bienaymé-Tchebychev inequality, we have:

$$\forall \epsilon > 0, \mathbb{P}_{\theta^2} (|X_N^{ik} - \mathbb{E}_{\theta^2} [X_N^{ik}]| \geq \epsilon) \leq \frac{\text{Var}_{\theta^2}(X_N^{ik})}{\epsilon^2}.$$

We have $\mathbb{E}_{\theta^2} [X_N^{ik}]$ which asymptotically behaves like $((\sigma_i^2)^2 \mathbb{1}_{k=i} + \sigma_i^2 \sigma_k^2 - \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}] \mathbb{E}_{\theta^2} [\epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}]])$ which is independent of N . Moreover, by changing ϵ in $\frac{\epsilon}{M^2}$, we have:

$$\forall \epsilon > 0, \mathbb{P}_{\theta^2} (M^2 |X_N^{ik} - \mathbb{E}_{\theta^2} [X_N^{ik}]| \geq \epsilon) \leq M^4 \frac{\text{Var}_{\theta^2}(X_N^{ik})}{\epsilon^2}.$$

Using assumption (A6), we have :

$$\text{Var}_{\theta^2} (X_N^{ik}) = \frac{1}{N^4} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \text{Var}_{\theta^2} (\text{Cov}_{\theta^2} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y}))$$

and, because the ϵ_{ij} are identically distributed for a fixed i , we have :

$$\mathbb{E}_{\theta^2} [(X_N^{ik})^2] \leq \frac{n_i n_k}{N^4} \mathbb{E}_{\theta^2} [(\text{Cov}_{\theta^2} (\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}), \epsilon_{k1} (\tau - \mathbb{1}_{\epsilon_{k1} \leq 0}) | \mathbf{Y}))^2]$$

Finally:

$$\forall \epsilon > 0, \mathbb{P}_{\theta^2} (|X_N^{ik} - \mathbb{E}_{\theta^2} [X_N^{ik}]| \geq \epsilon) \leq \frac{M^4 C}{N^2 \epsilon^2}$$

with C an unknown constant of N so we have , by construction of M , the convergence to 0.

Moreover, we always have:

$$\frac{1}{N^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}) | \mathbf{Y}] \xrightarrow{N \rightarrow +\infty} 0.$$

There is one last term left in the sum:

$$-\frac{1}{N^2} \frac{2}{(\sigma_i^2)^3} \mathbb{E}_{\theta^2} [\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0})] = -\frac{2}{N^2 (\sigma_i^2)^2}.$$

Consequently, as $\frac{M}{N} \xrightarrow{N \rightarrow +\infty} 0$, we have $M^2 \Phi_2 \xrightarrow{N \rightarrow +\infty} 0$.

Φ_3 :

We have Φ_3 which is a matrix of size $n \times n$ whose elements i, k are:

$$\begin{aligned} & \frac{1}{N^2} \left(\left(\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^0)^2} \right) + \frac{2\mathbb{1}_{k=i}}{(\sigma_i^2)^3} \mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})] - \frac{2\mathbb{1}_{k=i}}{(\sigma_i^0)^3} \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]] \right. \\ & \left. + \frac{n_i n_k}{(\sigma_k^2)^2} \mathbb{1}_{k=i} - \frac{1}{(\sigma_i^0 \sigma_k^0)^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \mathbb{E}_{\theta_2} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})] \right). \end{aligned}$$

As $\theta^2 \in N_N(\theta^0)$, we have :

$$\mathbb{E}_{\theta^2} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})] \xrightarrow{N \rightarrow +\infty} \mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})] = \sigma_i^0,$$

$$\mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]] \xrightarrow{N \rightarrow +\infty} \mathbb{E}_{\theta^0} [\mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]] = \mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})] = \sigma_i^0,$$

$$\mathbb{E}_{\theta_2} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})] \xrightarrow{N \rightarrow +\infty}$$

$$\mathbb{E}_{\theta_0} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})]$$

and

$$\mathbb{E}_{\theta_0} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})] = (\sigma_i^0)^2 \mathbb{1}_{k=i}.$$

We finally obtain Φ_3 which is equivalent to:

$$\frac{3 + n_i n_k}{N^2} \left(\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^0)^2} \right)$$

We saw before that:

$$\frac{1}{(\sigma_i^2)^2} - \frac{1}{(\sigma_i^0)^2} \leq \frac{M}{N} C$$

with C an unknown constant as $\theta^2 \in N_N(\theta^0)$.

As a consequence, because $\frac{M}{N} \xrightarrow{N \rightarrow +\infty} 0$ and $\frac{M^3}{N} \xrightarrow{N \rightarrow +\infty} 0$, we have $M^2 \Phi_3 \xrightarrow{N \rightarrow +\infty} 0$.

$\underline{\Phi}_4$:

We have Φ_4 which is a matrix of size $n \times n$ whose elements i, k are:

$$\frac{1}{N^2} \left(\frac{2}{(\sigma_i^0)^3} (\mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0}) | \mathbf{Y}]] - \mathbb{E}_{\theta^0} [\epsilon_{i1} (\tau - \mathbb{1}_{\epsilon_{i1} \leq 0})]) - \frac{n_i n_k}{(\sigma_k^0)^2} \mathbb{1}_{k=i} + \frac{1}{(\sigma_i^0 \sigma_k^0)^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} \mathbb{E}_{\theta^2} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\tau - \mathbb{1}_{\epsilon_{ij} \leq 0}), \epsilon_{kl} (\tau - \mathbb{1}_{\epsilon_{kl} \leq 0}) | \mathbf{Y})] \right).$$

We can see that the calculations and proofs of convergence are the same as for Φ_3 .

Finally, we've proved the hypothesis in the case (σ, σ) .

Determination for (σ, β)

$\underline{\Phi}_1$:

We have Φ_1 which is a matrix of size $n \times p$ whose elements i, k are:

$$\sup_{\theta^1 \in N_n(\theta^0)} \left\{ \frac{1}{N^2} \left(\frac{1}{(\sigma_i^1)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^1} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}] - \frac{1}{(\sigma_i^2)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}] + \frac{1}{(\sigma_i^1)^2} \sum_{p=1}^n \frac{1}{(\sigma_p^1)^2} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \text{Cov}_{\theta^1} (\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y}) - \frac{1}{(\sigma_i^2)^2} \sum_{p=1}^n \frac{1}{(\sigma_p^2)^2} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \text{Cov}_{\theta^2} (\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y}) \right) \right\}.$$

We still have, for all θ and by definition of M :

$$\frac{M^2}{N} \frac{1}{N} \sum_{j=1}^{n_i} \mathbb{E}_{\theta} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}]$$

which converge to 0.

Moreover, the difference between the third and fourth terms defining Φ_1 tends to :

$$\sum_{p=1}^n \left(\frac{1}{\sigma_i^1 \sigma_p^1} - \frac{1}{\sigma_i^2 \sigma_p^2} \right) \tau C_{pk} + \sum_{p=1}^n \left(\frac{1}{(\sigma_i^1)^2 \sigma_p^1} - \frac{1}{(\sigma_i^2)^2 \sigma_p^2} \right) \mathbb{E}_{\theta^0} [\mathbb{E}_{\theta^0} [\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) | \mathbf{Y}]] \mathbb{E}_{\theta^0} [\mathbb{1}_{\epsilon_{pq} \leq 0} | \mathbf{Y}]].$$

As before, it can be shown that the terms:

$$\frac{1}{\sigma_i^1 \sigma_p^1} - \frac{1}{\sigma_i^2 \sigma_p^2} \quad \text{and} \quad \frac{1}{(\sigma_i^1)^2 \sigma_p^1} - \frac{1}{(\sigma_i^2)^2 \sigma_p^2}$$

can be written as the difference between one or more parameters of θ_1 and θ_2 multiplied by a constant, so these terms are dominated by $\frac{M}{N}$.

As this is true for all $\theta^1 \in N_N(\theta^0)$, we can proceed to the sup and we have $M^2 \Phi_1 \xrightarrow{N \rightarrow +\infty} 0$.

Φ_2 :

We have Φ_2 which is a matrix of size $n \times p$ whose elements i, k are:

$$\begin{aligned} & \frac{1}{N^2} \left(\frac{1}{(\sigma_i^2)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}] \right. \\ & + \frac{1}{(\sigma_i^2)^2} \sum_{p=1}^n \frac{1}{\sigma_p^2} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \text{Cov}_{\theta^2} (\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y}) \\ & \left. - \frac{n_i}{(\sigma_i^2)^2} \sum_{p=1}^n \frac{1}{\sigma_p^2} \sum_{q=1}^{n_p} \tau \sigma_i^2 X_{pqk} - \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^2} [\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) | \mathbf{Y}] \mathbb{E}_{\theta^2} [\mathbb{1}_{\epsilon_{pq} \leq 0} | \mathbf{Y}]] X_{pqk} \right). \end{aligned}$$

We have, exactly like Φ_1 :

$$\frac{M^2}{N^2} \frac{1}{(\sigma_i^2)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}] \xrightarrow{N \rightarrow +\infty} 0.$$

The last term of the sum defining Φ_2 tends to:

$$-\frac{2\tau}{\sigma_i^2} \sum_{p=1}^n \frac{c_{pk}}{\sigma_p^2}$$

which is exactly the opposite of the limit of the second term of the sum defining Φ_2 . In the same way as in the calculation of Φ_2 for the case (σ, σ) , we can identify a difference between a random variable and its expectation. We can then apply Bienaymé-Tchebychev and obtain the same convergence result.

Φ_3 :

We have Φ_3 which is a matrix of size $n \times p$ whose elements i, k are:

$$\begin{aligned} & \frac{1}{N^2} \left(-\frac{1}{(\sigma_i^0)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} [\mathbb{E}_{\theta^0} [(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau) X_{ijk} | \mathbf{Y}]] + \frac{2\tau n_i}{\sigma_i^2} \sum_{p=1}^n \frac{1}{\sigma_p^2} \sum_{q=1}^{n_p} X_{pqk} \right. \\ & \left. - \frac{1}{(\sigma_i^0)^2} \sum_{p=1}^n \frac{1}{\sigma_p^0} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \mathbb{E}_{\theta^2} [\text{Cov}_{\theta^0} (\epsilon_{ij} (\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} | \mathbf{Y})] \right). \end{aligned}$$

We have :

$$\mathbb{E}_{\theta^2} \left[\mathbb{E}_{\theta^0} \left[\left(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau \right) X_{ijk} \mid \mathbf{Y} \right] \right] \xrightarrow{N \rightarrow +\infty} \mathbb{E}_{\theta^0} \left[\mathbb{E}_{\theta^0} \left[\left(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau \right) X_{ijk} \mid \mathbf{Y} \right] \right] = 0$$

et

$$\mathbb{E}_{\theta^2} \left[\text{Cov}_{\theta^0} \left(\epsilon_{ij} \left(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau \right), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} \mid \mathbf{Y} \right) \right] \xrightarrow{N \rightarrow +\infty} 2\tau \sigma_i^0 X_{pqk}.$$

We then find that Φ_3 is equivalent to:

$$2\tau \sum_{p=1}^n c_{pk} \left(\frac{1}{\sigma_i^2 \sigma_p^2} - \frac{1}{\sigma_i^0 \sigma_p^0} \right).$$

We saw before that $\left| \frac{1}{\sigma_i^2 \sigma_p^2} - \frac{1}{\sigma_i^0 \sigma_p^0} \right| \leq C \frac{M}{N}$, and so we have $M^2 \Phi_3 \xrightarrow{N \rightarrow +\infty} 0$.

Φ_4 :

We have Φ_4 which is a matrix of size $n \times p$ whose elements i, k are:

$$\begin{aligned} & \frac{1}{N^2} \left(\frac{1}{(\sigma_i^0)^2} \sum_{j=1}^{n_i} \mathbb{E}_{\theta^2} \left[\mathbb{E}_{\theta^0} \left[\left(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau \right) X_{ijk} \mid \mathbf{Y} \right] \right] - \frac{2\tau n_i}{\sigma_i^0} \sum_{p=1}^n \frac{1}{\sigma_p^0} \sum_{q=1}^{n_p} X_{pqk} \right. \\ & \left. + \frac{1}{(\sigma_i^0)^2} \sum_{p=1}^n \frac{1}{\sigma_p^0} \sum_{j=1}^{n_i} \sum_{q=1}^{n_p} \mathbb{E}_{\theta^2} \left[\text{Cov}_{\theta^0} \left(\epsilon_{ij} \left(\mathbb{1}_{\epsilon_{ij} \leq 0} - \tau \right), \mathbb{1}_{\epsilon_{pq} \leq 0} X_{pqk} \mid \mathbf{Y} \right) \right] \right). \end{aligned}$$

We have the convergence result using the continuity of the expectation in θ^2 and the results used for the case of Φ_3 .

Finally, we have shown the hypothesis in the (β, σ) case.

Determination for (β, β)

Φ_1 :

Φ_1 can be expressed:

$$\begin{aligned} & \sup_{\theta^1 \in N_n(\theta^0)} \left\{ \frac{1}{N^2} \sum_{i=1}^n \sum_{k=1}^n \left(\frac{1}{\sigma_i^1 \sigma_k^1} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} X_{ij} \text{Cov}_{\theta^1} \left(\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y} \right) X_{kl}^t \right. \right. \\ & \left. \left. - \frac{1}{\sigma_i^2 \sigma_k^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} X_{ij} \text{Cov}_{\theta^2} \left(\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y} \right) X_{kl}^t \right) \right\}. \end{aligned}$$

As $\theta^1, \theta^2 \in N_N(\theta^0)$, we have:

$$\text{Cov}_{\theta} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y}) \xrightarrow{N \rightarrow +\infty} \text{Cov}_{\theta^0} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y})$$

for $\theta \in \{\theta^1, \theta^2\}$. Therefore, Φ_1 is equivalent to:

$$\sum_{i=1}^n \sum_{k=1}^n \left(\frac{1}{\sigma_i^1 \sigma_k^1} - \frac{1}{\sigma_i^2 \sigma_k^2} \right) C_{ik}^1 \text{Cov}_{\theta^0} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y}) X_{kl}^t.$$

As before, we then have convergence by passing to the sup, the term majorization:

$$\sum_{i=1}^n \sum_{k=1}^n \left(\frac{1}{\sigma_i^1 \sigma_k^1} - \frac{1}{\sigma_i^2 \sigma_k^2} \right)$$

and the (A5) assumption.

Φ_2 :

Φ_2 can be expressed:

$$\frac{1}{N^2} \sum_{i=1}^n \sum_{k=1}^n \frac{1}{\sigma_i^2 \sigma_k^2} \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} (X_{ij} \text{Cov}_{\theta^2} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y}) X_{kl}^t - X_{ij} (\tau - \tau^2) X_{kl}^t \mathbb{1}_{k=i}).$$

As before, we identify the difference between a random variable and its expectation.

Φ_3 :

Φ_3 is equal to:

$$\frac{1}{N^2} \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} X_{ij} \left(\frac{\tau - \tau^2}{\sigma_i^2 \sigma_k^2} \mathbb{1}_{k=i} - \frac{\mathbb{E}_{\theta^2} [\text{Cov}_{\theta^0} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y})]}{\sigma_i^0 \sigma_k^0} \right) X_{kl}^t.$$

The following result of convergence

$$\mathbb{E}_{\theta^2} [\text{Cov}_{\theta^0} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} \mid \mathbf{Y})] \xrightarrow{N \rightarrow +\infty} (\tau - \tau^2) \mathbb{1}_{k=i}$$

and the inequality

$$\sum_{i=1}^n \sum_{k=1}^n \left(\frac{1}{\sigma_i^2 \sigma_k^2} - \frac{1}{\sigma_i^0 \sigma_k^0} \right) \leq C \frac{M}{N}$$

allow us to conclude on the good convergence behavior of $M^2 \Phi_3$.

Φ_4 :

Φ_4 is written:

$$\frac{1}{N^2} \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^{n_i} \sum_{l=1}^{n_k} X_{ij} \left(\frac{\mathbb{E}_{\theta^0} [\text{Cov}_{\theta^0} (\mathbb{1}_{\epsilon_{ij} \leq 0}, \mathbb{1}_{\epsilon_{kl} \leq 0} | \mathbf{Y})]}{\sigma_i^0 \sigma_k^0} - \frac{\tau - \tau^2}{\sigma_i^0 \sigma_k^0} \mathbb{1}_{k=i} \right) X_{kl}^t.$$

Arguments similar to the case of Φ_3 allow us to conclude on the convergence properties of Φ_4 .

Finally, the conditions of the Theorem 3 of Weiss are satisfied.

Supplementary Figures : Asymptotic normality and convergence

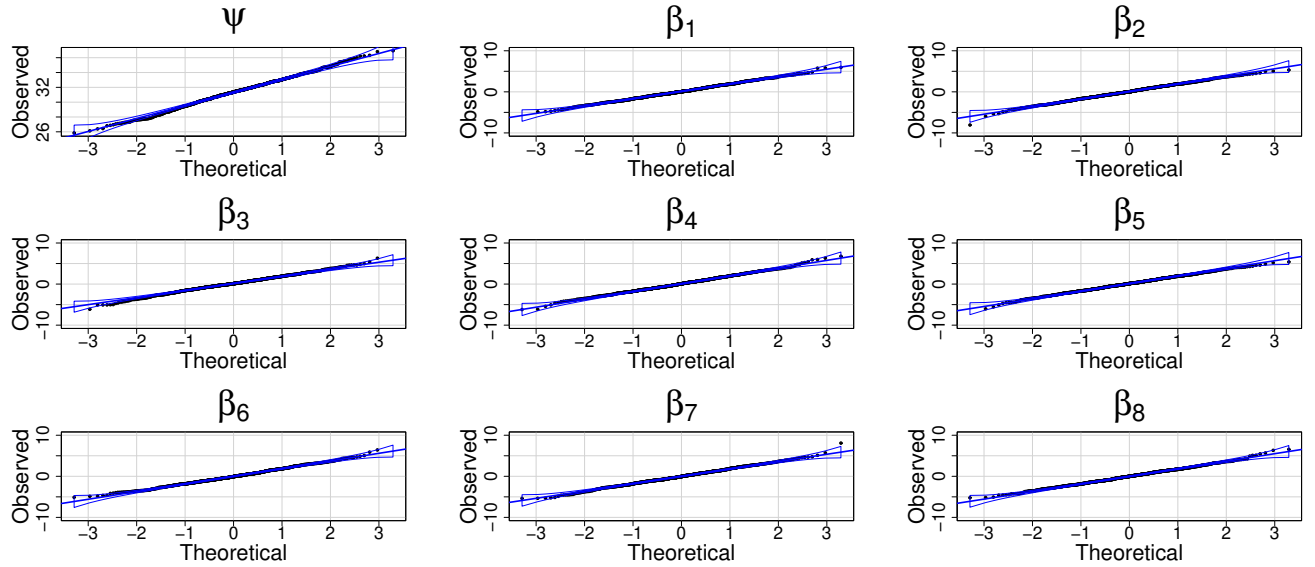


Figure 5: QQ-plot of $\hat{\beta} - \beta^0$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$.

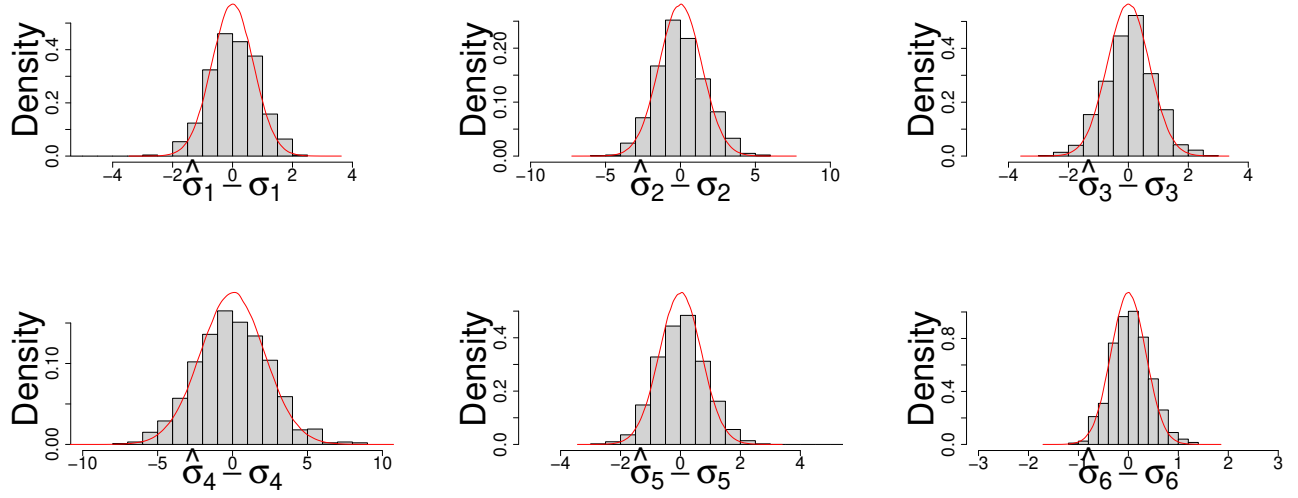


Figure 6: Empirical distribution of $\hat{\sigma} - \sigma$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$.

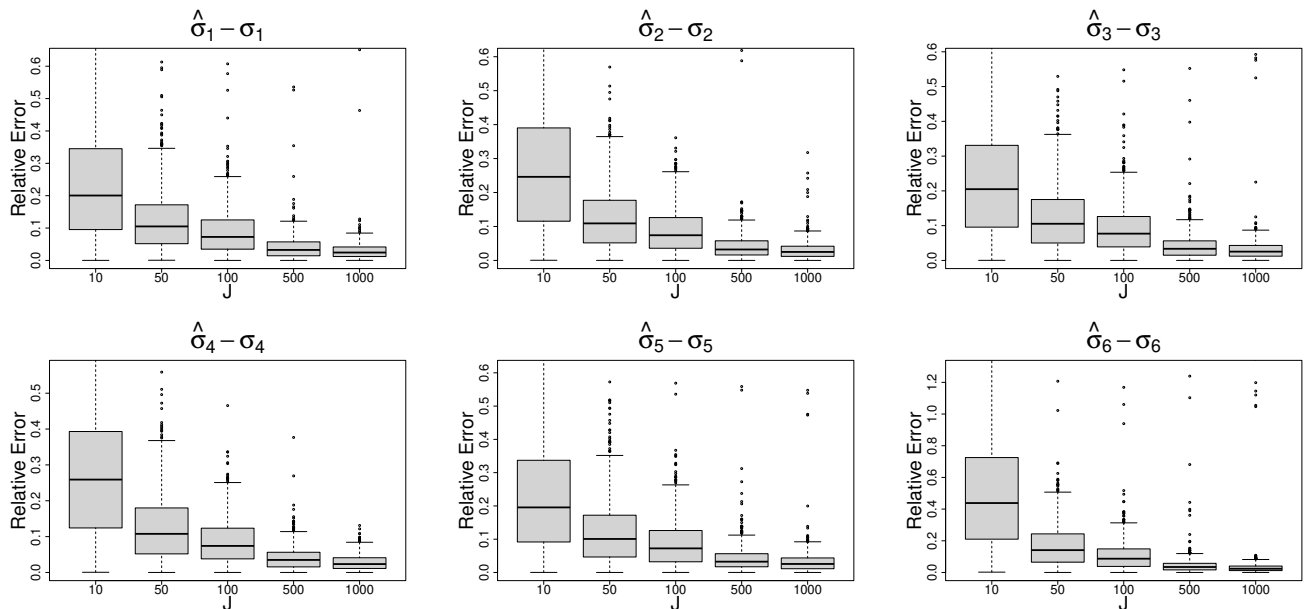


Figure 7: Convergence plot of $\hat{\sigma} - \sigma^0$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$.

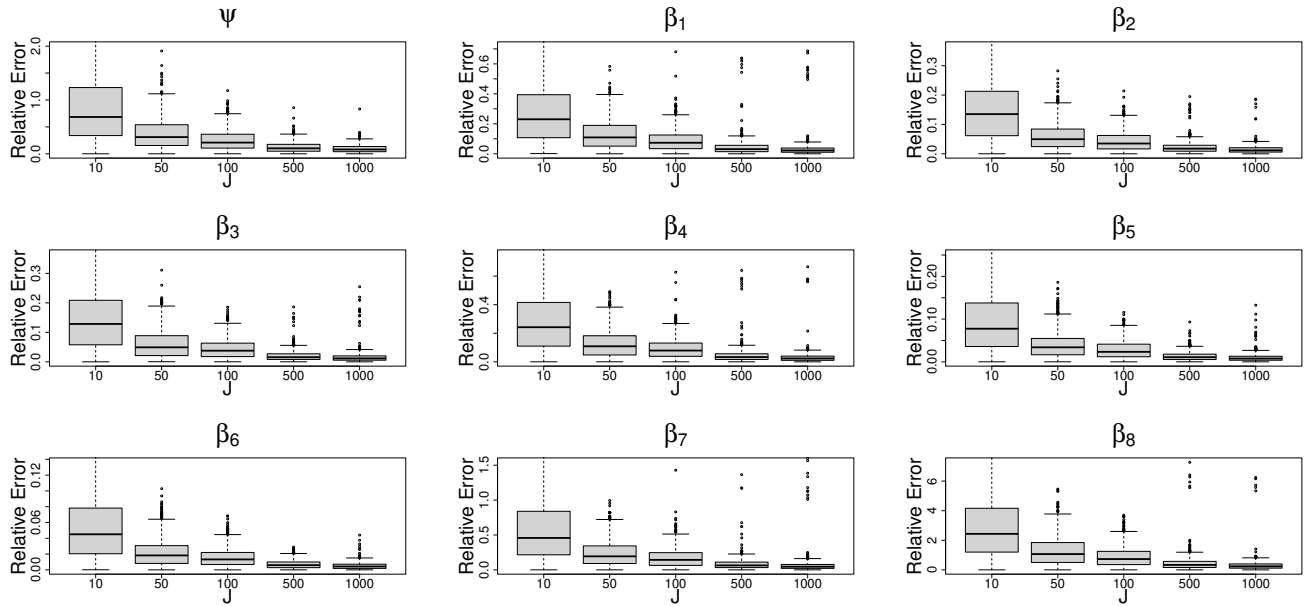


Figure 8: Convergence plot of $\hat{\beta} - \beta^0$ for a normal generation of co-variables following $\mathcal{N}(2, 2)$.

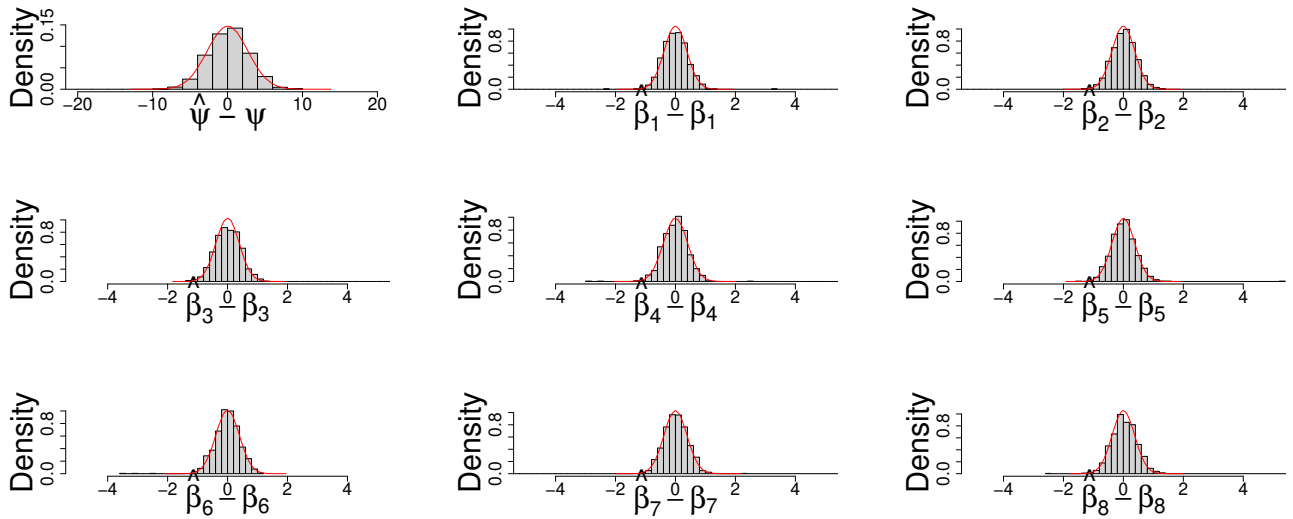


Figure 9: Empirical distribution of $\hat{\beta} - \beta$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$. For the estimator $\hat{\psi}$, the empirical bias was removed.

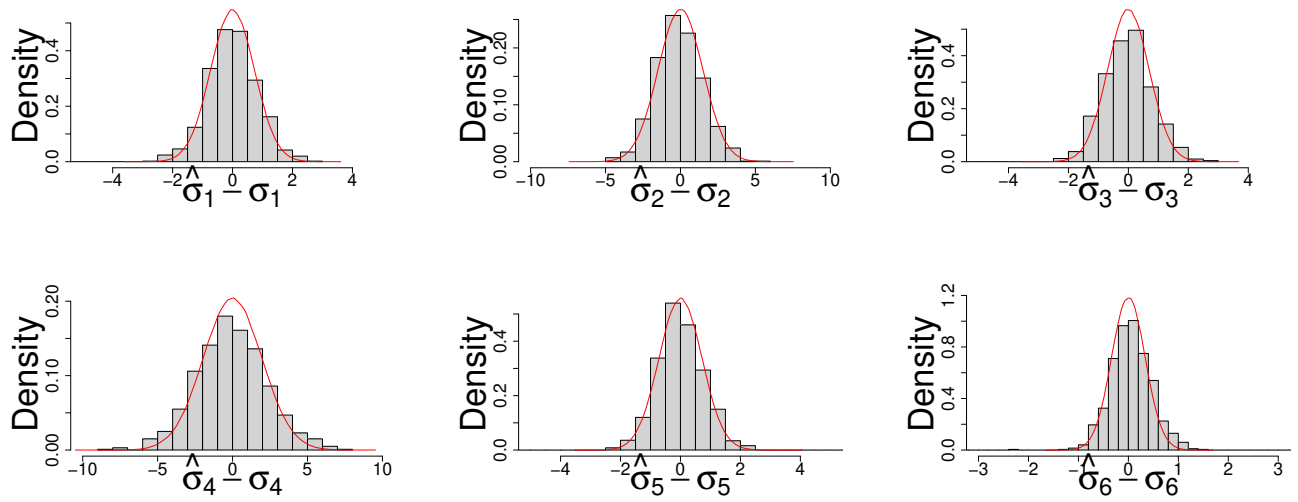


Figure 10: Empirical distribution of $\hat{\sigma} - \sigma$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$.

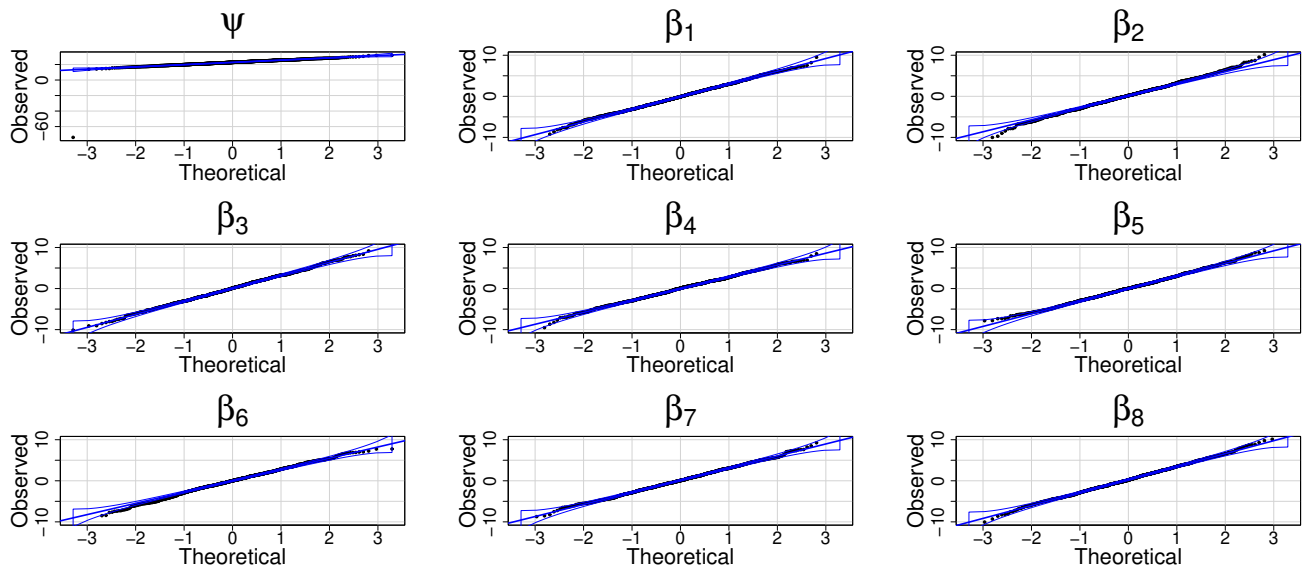


Figure 11: QQ-plot of $\hat{\beta} - \beta^0$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$.

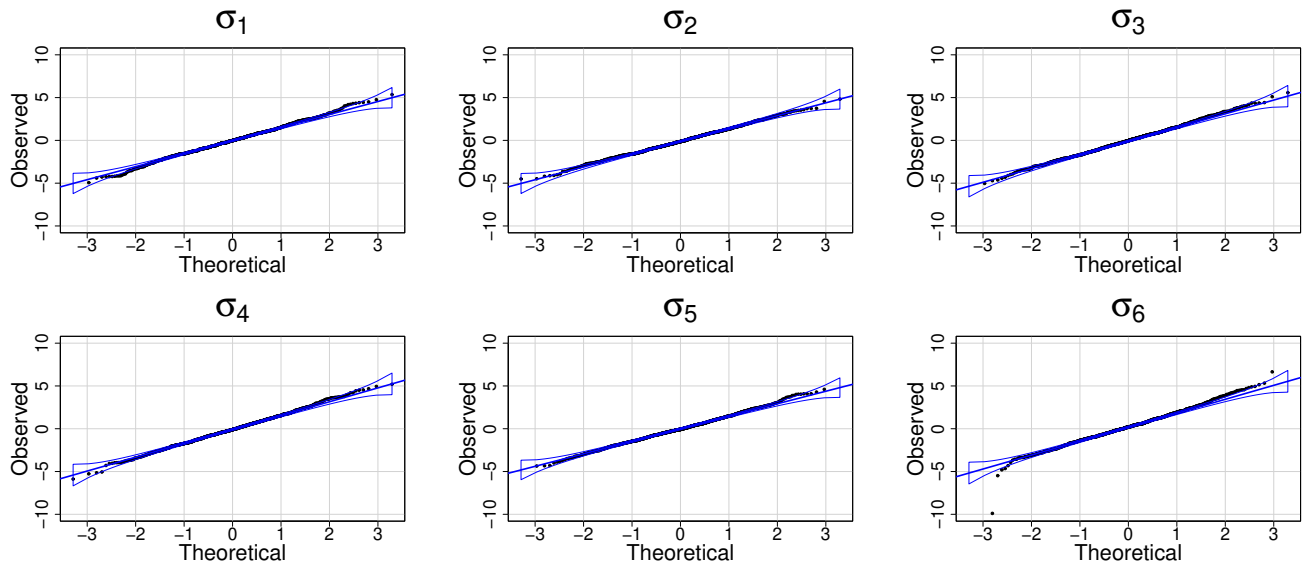


Figure 12: QQ-plot of $\hat{\sigma} - \sigma^0$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$.

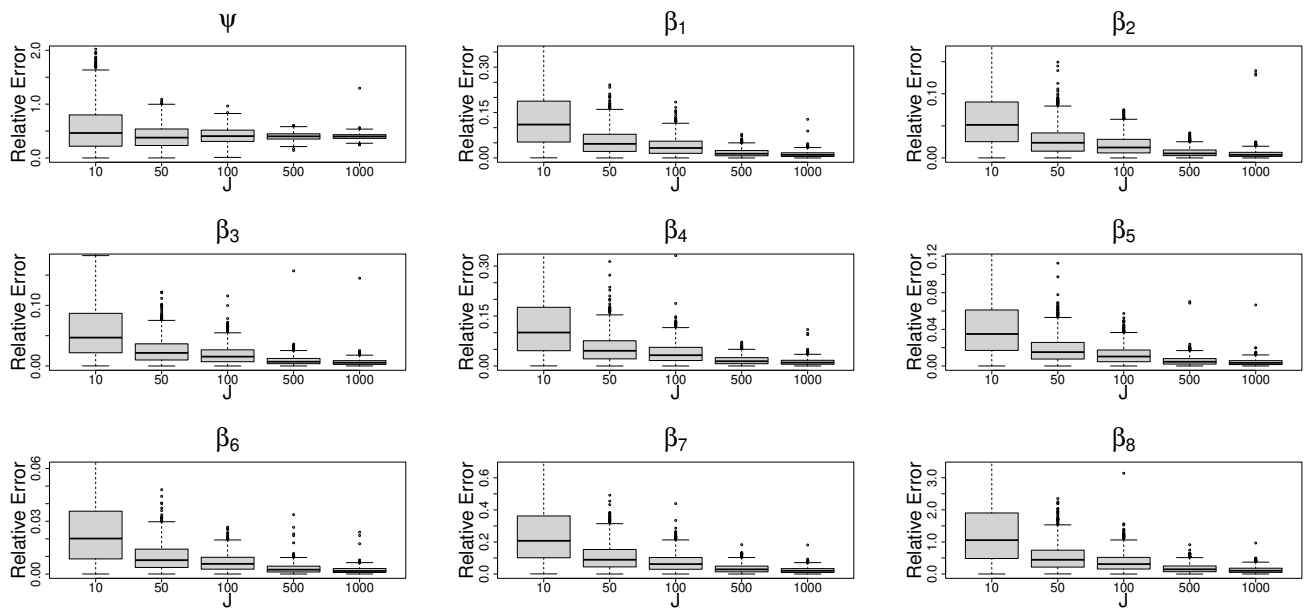


Figure 13: Convergence plot of $\hat{\beta} - \beta^0$ for a Laplace generation of co-variables following $\mathcal{L}(2, 3)$.

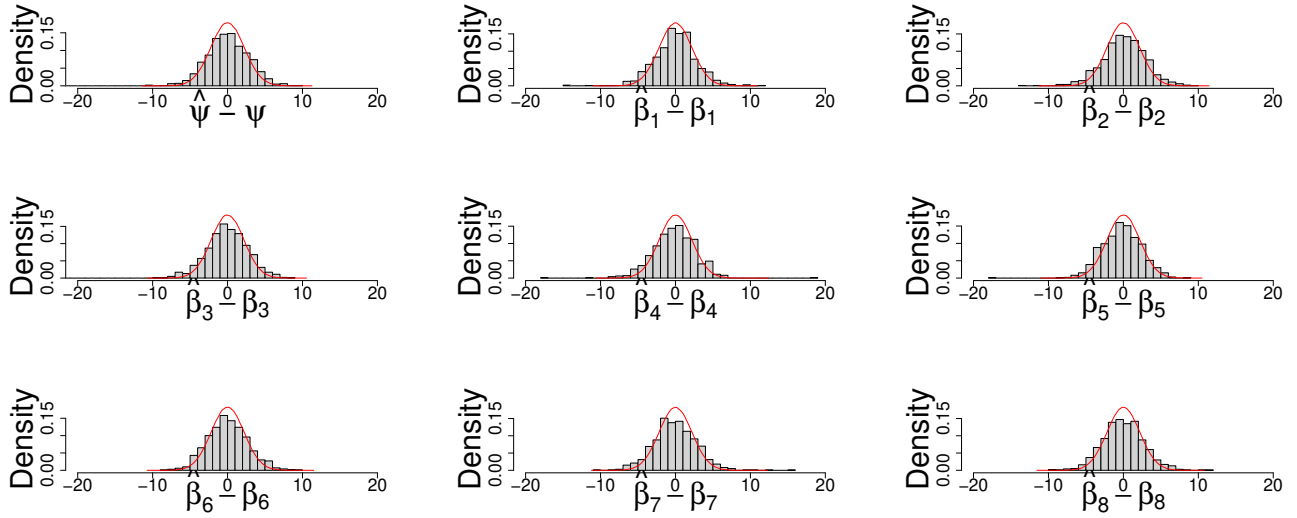


Figure 14: Empirical distribution of $\hat{\beta} - \beta$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$. For the estimator $\hat{\psi}$, the empirical bias was removed.

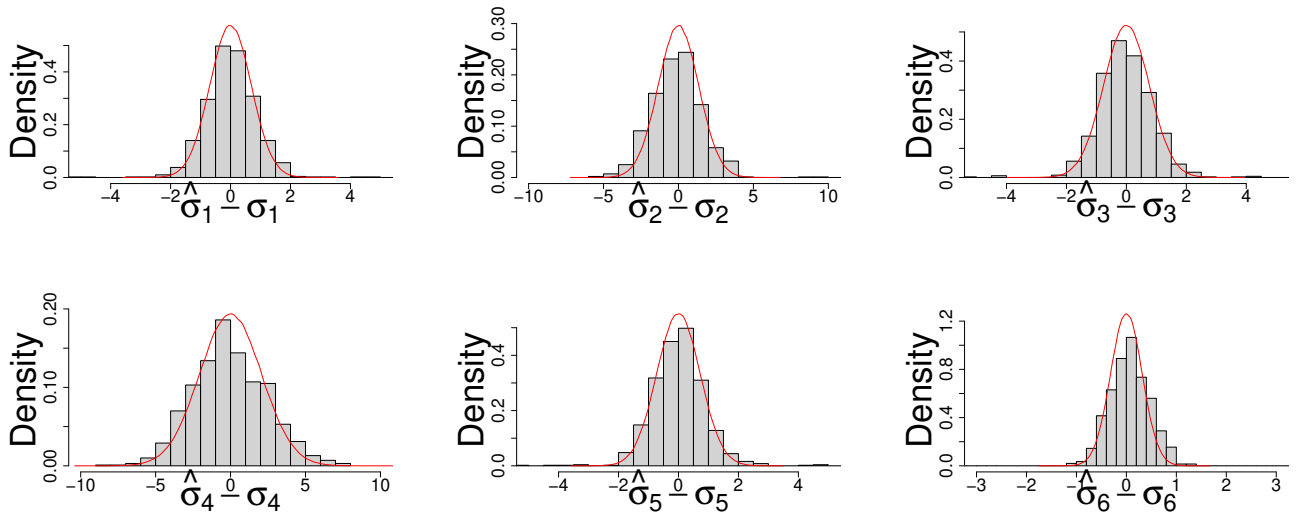


Figure 15: Empirical distribution of $\hat{\sigma} - \sigma$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$.

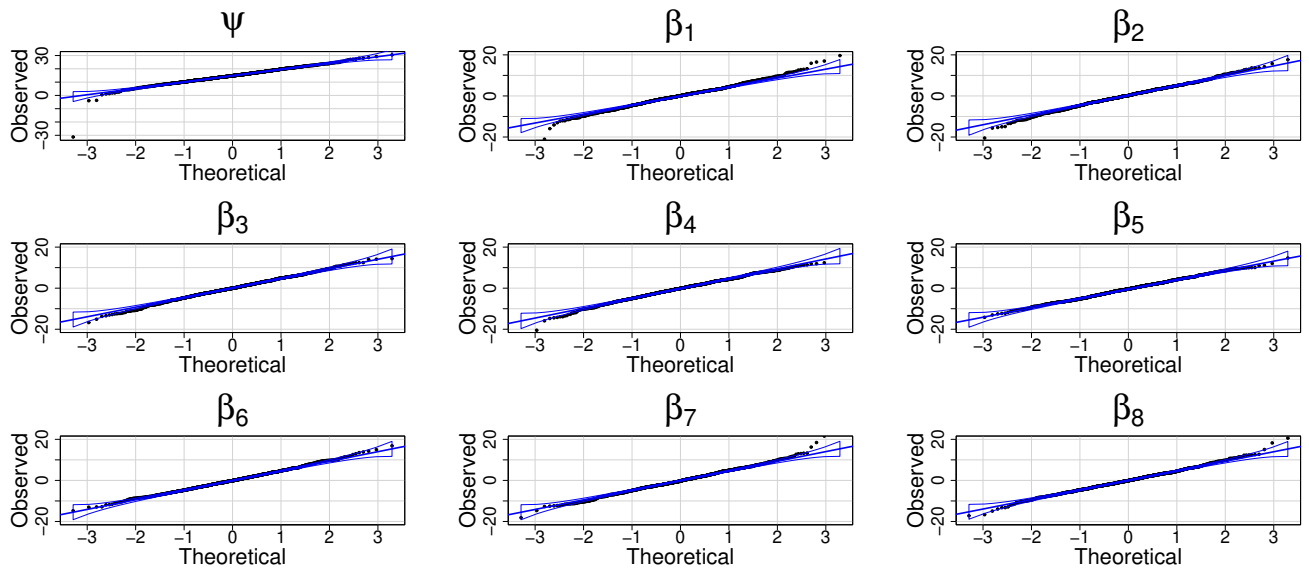


Figure 16: QQ-plot of $\hat{\beta} - \beta^0$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$.

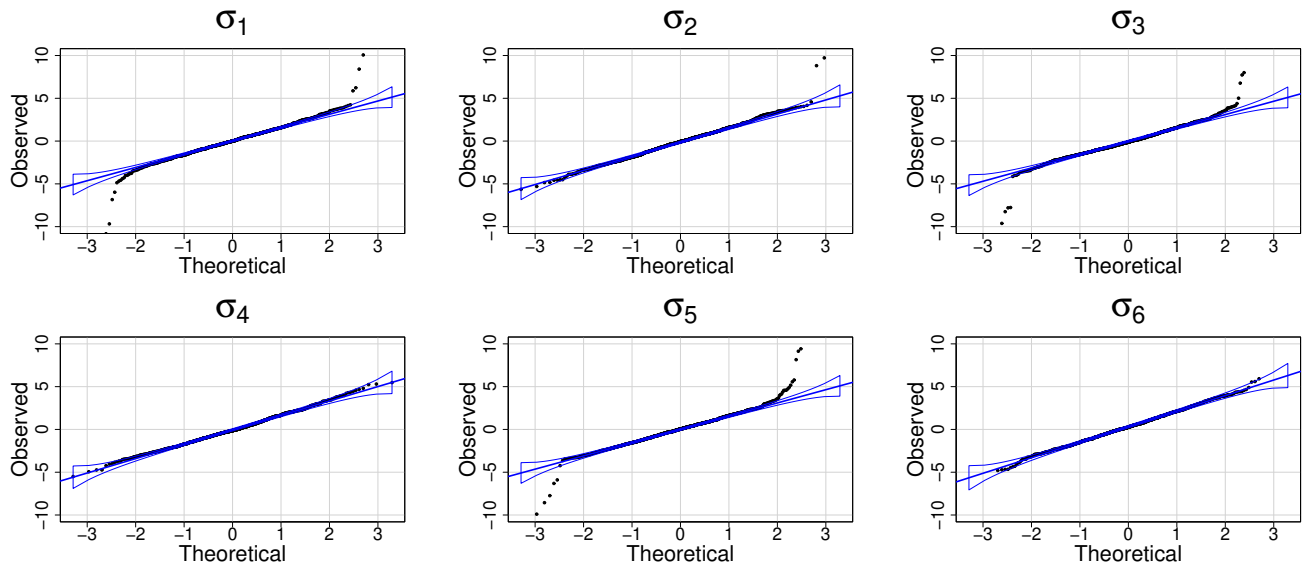


Figure 17: QQ-plot of $\hat{\sigma} - \sigma^0$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$.

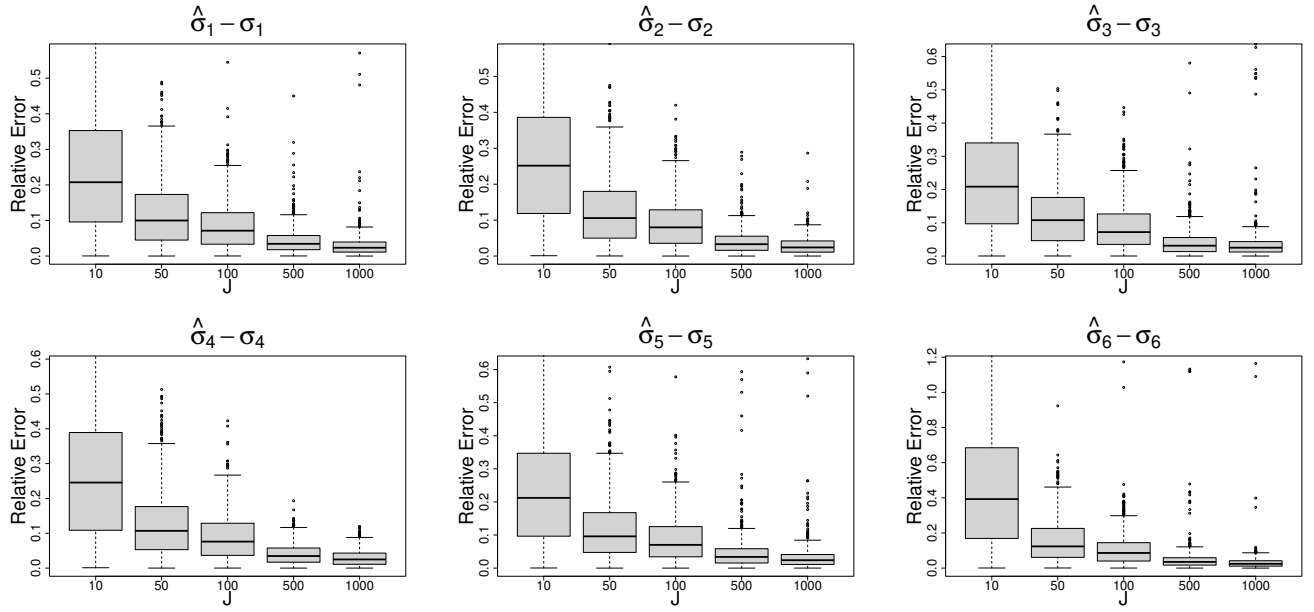


Figure 18: Convergence plot of $\hat{\sigma} - \sigma^0$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$.

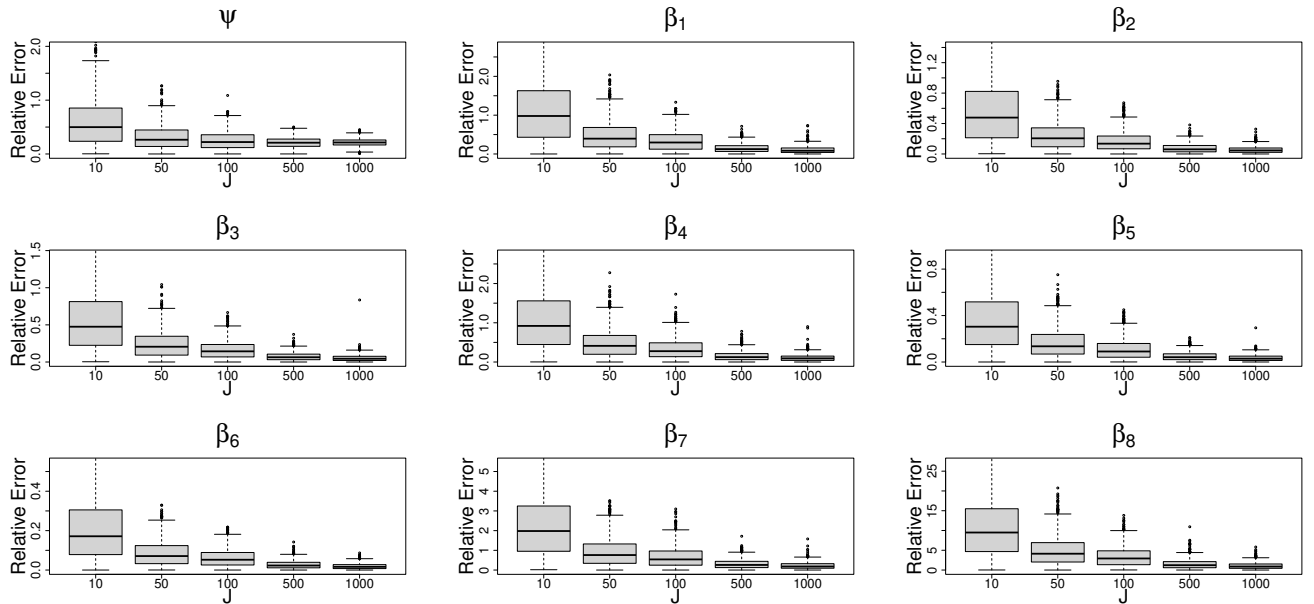


Figure 19: Convergence plot of $\hat{\beta} - \beta^0$ for a Bernoulli generation of co-variables following $\mathcal{B}(0.33)$.