

# RÉGRESSION QUANTILE PÉNALISÉE POUR DES DONNÉES LONGITUDINALES AVEC HÉTÉROSCÉDASTICITÉ.

Angelo Alcaraz<sup>1</sup>, Audrey poterie<sup>2</sup> & Gilles Durrieu<sup>3</sup>

<sup>1</sup> *Univ Bretagne Sud, CNRS UMR 6205, LMBA, France , [angelo.alcaraz@univ-ubs.fr](mailto:angelo.alcaraz@univ-ubs.fr)*

<sup>2</sup> *Univ Bretagne Sud, CNRS UMR 6205, LMBA, France , [audrey.poterie@univ-ubs.fr](mailto:audrey.poterie@univ-ubs.fr)*

<sup>3</sup> *Univ Bretagne Sud, CNRS UMR 6205, LMBA, France , [gilles.durrieu@univ-ubs.fr](mailto:gilles.durrieu@univ-ubs.fr)*

**Résumé.** La présence d'hétéroscédasticité dans les données peut souvent mettre en difficulté un processus de modélisation statistique. Dans ce contexte, les modèles mixtes et modèles pour données longitudinales ont été développés. Dans le contexte des modèles mixtes et des données longitudinales, cet article aborde directement ce problème. Plus précisément, notre objectif dans ce travail est de permettre d'améliorer l'estimation de la dispersion des données dans le cadre de données hétéroscédasticités. Pour ce faire, nous développons un nouvel estimateur quantile basé sur la distribution asymétrique de Laplace, qui explique l'hétéroscédasticité entre différents groupes d'individus. Outre le développement de ce nouveau modèle, ce travail établit aussi les bonnes propriétés asymptotiques de cet estimateur sous des hypothèses minimales sur les données et les vérifient à l'aide de simulations. En utilisant le formalisme permissif de la distribution de Laplace asymétrique, nous démontrons les propriétés asymptotiques d'une classe d'estimateurs définis par un problème d'optimisation généralisé inspiré du maximum de vraisemblance. Une pénalisation Ridge est aussi proposée pour traiter les problèmes de surestimation de la variabilité. Plus généralement, cet article présente un modèle permettant de traiter plus précisément les problèmes d'estimation de volume. La méthode a été implémentée en R, l'ensemble des fonctions est disponible sur Github.

**Mots-clés.** Distribution de Laplace asymétrique, modèles linéaires mixtes quantiles, quadrature gaussienne, régression Ridge.

**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. The method has been implemented in R, and the full set of functions is available on Github.

**Keywords.** Asymmetric Laplace Distribution, Linear Quantile Mixed Models, Gaussian Quadrature, Ridge Regression.

## 1 Introduction

Dans cet article, nous proposons une théorie formelle pour traiter l'hétéroscédasticité dans un modèle de données de panel. Dans certaines applications, il peut être utile de disposer d'un estimateur précis de la dispersion des données, en particulier pour mesurer l'hétéroscédasticité. C'est ce que tente de faire le modèle présenté dans cet article, en mentionnant explicitement les différences de variabilité entre différents groupes de données. De plus, le formalisme mathématique sous-jacent proposé permet d'ajouter simplement des pénalités, ce qui permet de résoudre les problèmes de surestimation de la variance. En étudiant plusieurs variables indépendantes d'intérêt à l'aide du modèle présenté dans cet article, il est alors possible d'effectuer une estimation de volume au sens large (pour des espaces de dimensions supérieures à 1), en contrôlant la dispersion intra-groupe de chacune de ces variables. Considérons le modèle suivant:

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

où pour tout  $n \geq 1$ ,  $\mathbf{Y}_n = (Y_1, \dots, Y_n)^t$  est le vecteur des observations,  $\mathbf{X}_n$  est une matrice connue de dimension  $n \times p$  de lignes  $\mathbf{x}_i^t \in \mathbb{R}^p$ ,  $i = 1, \dots, n$ ,  $\boldsymbol{\epsilon}_n = (\epsilon_1, \dots, \epsilon_n)^t$  est un vecteur d'erreur i.i.d et  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$  définit le vecteur de paramètres inconnus à estimer. On appelle  $\tau$ -ième quantile de régression (Koenker and Bassett (1978)),  $0 < \tau < 1$ , toute solution du problème de minimisation :

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

ou  $\rho_\tau(u) = u(\tau - \mathbb{1}_{u < 0})$  avec  $\mathbb{1}_{\mathcal{P}}$  qui prends les valeurs 1 ou 0 selon la réalisation ou non de la condition  $\mathcal{P}$ . La fonction  $\rho$  est appelée "fonction de perte du quantile" et est classiquement utilisée pour définir le quantile d'une variable aléatoire. Le développement de la régression quantile pour les données de panel suit en quelque sorte le chemin du développement des modèles linéaires généralisés (Liang and Zeger (1986)). Koenker (2004) s'est intéressé aux données longitudinales, une forme classique de données que l'on peut trouver en médecine, en écologie ou en biologie. Dans ce paradigme, nous tenons compte de la dépendance entre des mesures groupées (par exemple, une mesure médicale sur le même sujet). Ceci a conduit à la formulation d'un modèle de régression quantile à effets fixes, exprimé sous la forme d'une pénalité lasso (least absolute shrinkage and selection operator) (Lamarche (2010)).

Nous examinons ici le développement de la régression quantile basée sur la distribution de Laplace asymétrique (Yu and Zhang (2005)). Plusieurs articles fondamentaux ont établi le lien entre le problème de minimisation de la régression quantile et l'estimateur du maximum de vraisemblance (Geraci and Bottai (2007)). Cela a permis d'utiliser les algorithmes classiques de résolution de l'estimateur du maximum de vraisemblance pour

déterminer l'estimateur de la régression du quantile. On notera l'utilisation de la théorie bayésienne ( Aghamohammadi and Mohammadi (2017), Ji and Shi (2022)) ou de l'algorithme *Expectation-Maximisation* (Geraci and Bottai (2014), Geraci (2014), Galarza et al. (2017) Tian et al. (2020), Battagliola et al. (2022)). Cette communication suit les traces de ces auteurs, en profitant de la distribution de Laplace asymétrique et de son lien avec la régression par quantile et en l'utilisant sur des données longitudinales avec des effets aléatoires. Nous présentons dans cet article le modèle et quelques résultats théoriques de nos estimateurs, avant d'illustrer les performances de notre approche sur des simulations.

## 2 Modèle et résultats théoriques

### 2.1 Régression quantile pour des données longitudinales hétéroscédastiques utilisant la distribution asymétrique de Laplace

Nous considérons le cadre des données longitudinales. Soit  $n \in \mathbb{N}$  le nombre d'individus. L'indice  $i \in \{1, \dots, n\}$  dénote le niveau des individus et, pour chaque individu  $i$ , on considère  $n_i \in \mathbb{N}$  mesures. L'indice  $j \in \{1, \dots, n_i\}$  correspond à la  $j$ -ième mesure du  $i$ -ième individu. Soit  $\mathbf{Y} = (Y_{ij})_{1 \leq i \leq n, 1 \leq j \leq n_i} \in \mathbb{R}^N$  une variable réponse étudiée, avec  $N = \sum_{i=1}^n n_i$ . La forme générale du modèle mixte utilisé est la suivante :

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

où pour tous les  $i \in \{1, \dots, n\}$  et tous les  $j \in \{1, \dots, n_i\}$ ,  $\mathbf{X}_{ij} \in \mathbb{R}^p$  sont les observations pour la  $j$ -ième mesure du  $i$ -ième individu,  $p \in \mathbb{N}$  est le nombre de variables,  $\boldsymbol{\beta} \in \mathbb{R}^p$  sont les paramètres de la régression (qui peuvent contenir des effets fixes),  $e_i \in \mathbb{R}^n$  est le  $i$ -ième vecteur de la base standard,  $\boldsymbol{\nu} \in \mathbb{R}^n$  est le vecteur des effets aléatoires, décrivant un effet au niveau individuel comme une ordonnée à l'origine, et  $\epsilon_{ij}$  est une variable aléatoire centrée représentant le terme d'erreur. Suivant Geraci and Bottai (2014), nous ajoutons deux autres hypothèses sur le modèle. Premièrement, on prend  $\boldsymbol{\nu} \sim \mathcal{N}(0, \phi \Sigma)$ , avec  $\phi \in \mathbb{R}$  un paramètre à fixé et  $\Sigma \in \mathbb{R}^{n \times n}$  une matrice définie positive connue. Cette dernière est estimée plus loin et représentera la structure de dépendance entre les individus. Deuxièmement, nous supposons que  $\epsilon_{ij}$  suit une distribution de Laplace asymétrique, comme dans Geraci and Bottai (2007). Une variable aléatoire  $Z$  suit une distribution de Laplace asymétrique de paramètres  $(\mu, \sigma, \tau)$  ( qui est dénotée par  $Z \sim ALD(\mu, \sigma, \tau)$  ) si sa densité peut être exprimée comme :

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

avec  $\rho_\tau(u) = u(\tau - \mathbb{1}_{u < 0})$  désignant la fonction de perte quantile. Une étude détaillée de la distribution de Laplace asymétrique est proposée par Yu and Zhang (2005). Plus précisément, nous supposons que  $\epsilon_{ij} \sim ALD(0, \sigma_i, \tau)$ , où  $0 < \tau < 1$  représente l'ordre du quantile et  $\boldsymbol{\sigma} = (\sigma_i)_{1 \leq i \leq n} \in (\mathbb{R}_+^*)^n$  correspond aux paramètres d'échelle. En utilisant cette hypothèse, nous prenons explicitement en compte l'hétéroscédasticité qui existe entre les individus. De plus, nous supposons que tous les  $\epsilon_{ij}$  sont indépendants les uns des autres et des effets

aléatoires. Finalement, nous devons déduire le vecteur  $\boldsymbol{\theta} = (\boldsymbol{\sigma}, \boldsymbol{\beta}) \in (\mathbb{R}_+^*)^n \times \mathbb{R}^p$ . En ce qui concerne l'asymptotique, l'une des conséquences de ce choix de modélisation est la manière de tendre vers l'infini. Comme le nombre de paramètres à estimer dépend maintenant de  $n$ , nous ne nous intéressons pas à la limite lorsque  $n$  tend vers l'infini. Nous déterminons plutôt la limite de notre estimateur lorsque pour tout  $1 \leq i \leq n, n_i \rightarrow +\infty$ , c'est-à-dire lorsque le nombre de mesures par individu tend vers l'infini.

Nous définissons notre estimateur  $\hat{\boldsymbol{\theta}}$  comme un estimateur du maximum de vraisemblance de  $\boldsymbol{\theta}$ . Soit

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

la densité conditionnelle de la réponse par rapport au paramètres et  $\mu_{ij} = \mathbf{X}_{ij}^t \boldsymbol{\beta} + \mathbf{e}_i^t \mathbf{v}$ . Grâce à l'hypothèse d'indépendance on obtient pour tout  $y$  et  $\mathbf{v}$  la densité multivariée

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

En intégrant (4) sur les effets aléatoires, on obtient la vraisemblance. On a d'abord :

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

Il en découle que l'estimateur  $\hat{\boldsymbol{\theta}}$  est défini par

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

avec  $\Theta \subset (\mathbb{R}_+^*)^n \times \mathbb{R}^p$  l'espace des paramètres. Le problème de maximisation est équivalent au problème de minimisation de l'estimateur de la régression quantile (Koenker and Bassett (1978)). Cette équivalence est la principale motivation de l'utilisation de la distribution de Laplace asymétrique. En outre, nous définissons un autre estimateur  $\hat{\boldsymbol{\theta}}_f$  de  $\boldsymbol{\theta}$  qui est défini comme la solution du problème de maximisation suivant

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

où  $(f_{n_i})_{1 \leq i \leq n}$  est une suite de fonctions prenant ses valeurs dans l'espace des paramètres  $\Theta$ . Ce problème d'optimisation correspond au problème d'optimisation du maximum de vraisemblance avec un terme de pénalité  $\sum_{i=1}^n f_{n_i}(\boldsymbol{\theta})$  supplémentaire. Ce terme de régularisation peut être ajouté au problème d'optimisation de maximum de vraisemblance dans le but par exemple de réduire la complexité du modèle et d'éviter l'*overfitting* (Tibshirani (1996), Friedman et al. (2000)).

## 2.2 Résultats théoriques

Afin d'étudier le comportement asymptotique de nos estimations, il est nécessaire d'introduire plusieurs hypothèses.

- A1.** On a  $N \rightarrow +\infty$  et pour tout  $i \in \{1, \dots, n\}$ ,  $n_i \rightarrow +\infty$  tel que  $n_i = \mathcal{O}(N)$ .
- A2.** La vraie valeur de  $\boldsymbol{\theta}$  est notée  $\boldsymbol{\theta}^0$  et on a  $\boldsymbol{\theta}^0 \in \text{int}(\Theta)$ , où  $\text{int}(\Theta)$  est l'intérieur de  $\Theta$ .
- A3.** Pour tout  $i \in \{1, \dots, n\}$  et  $j \in \{1, \dots, n_i\}$ , les variables aléatoires  $\epsilon_{ij}$  sont mutuellement indépendantes et indépendantes de  $\nu$ .
- A4.** La suite  $\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  tends vers  $c_i \in \mathbb{R}^p$  où  $n_i \rightarrow +\infty$  pour tout  $i \in \{1, \dots, n\}$ .
- A5.** La suite  $\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} X_{ij}^t$  tend vers une matrice définie positive  $C_i \in \mathbb{R}^{p \times p}$  où  $n_i \rightarrow +\infty$  for all  $i \in \{1, \dots, n\}$ .
- A6.** Pour tout  $i \in \{1, \dots, n\}$  et  $j \in \{1, \dots, n_i\}$ , les variables aléatoires  $\mathbb{E}[\epsilon_{ij} | \mathbf{Y}]$  sont mutuellement indépendantes et indépendantes de  $\mathbb{E}[\nu | \mathbf{Y}]$ .

L'hypothèse **(A1)** permet de définir la façon dont notre modèle va atteindre l'infini, comme explicité dans la partie 2.1. L'hypothèse **(A2)** est une hypothèse classique pour la régression quantile et est nécessaire à l'application du théorème principal utilisé lors de la preuve du comportement asymptotique de notre estimateur. Les hypothèses **(A3)** à **(A5)** limitent la forme que peuvent prendre les données et garantissent l'existence de moments du premier et du second ordre. Ces hypothèses sont également couramment utilisées dans la régression quantile. Nous pouvons constater que ni les données ni le terme de variabilité  $\epsilon_{ij}$  n'ont besoin de suivre une distribution normale pour que les résultats asymptotiques soient établis. La dernière hypothèse **(A6)** est nécessaire pour la preuve du Théorème 1 et est rarement utilisée dans la littérature. On peut voir une utilisation et une explication détaillée de cette hypothèse dans Weidenhammer (2017). Les principaux résultats de ce document sont les théorèmes suivants qui établissent la normalité asymptotique de nos estimateurs.

**Théorème 1** *Sous les hypothèses (A1) à (A6), on a, lorsque  $N$  tends vers l'infini, la normalité asymptotique suivante*

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

où  $B(\boldsymbol{\theta})$  est une matrice définie positive de taille  $(n + p) \times (n + p)$ .

De plus, nous pouvons calculer la valeur de  $B(\boldsymbol{\theta}^0)$ , ce qui nous permet d'établir des intervalles de confiance précis. Cela nous permet également de comparer la distribution asymptotique du paramètre avec celle de l'estimateur et nous renseigne sur la vitesse de convergence en  $\sqrt{N}$ . La cohérence de notre estimateur découlera immédiatement de ce résultat, comme dans Weidenhammer (2017). Nous définissons deux autres hypothèses sur  $(f_{n_i})_{1 \leq i \leq n}$  de l'équation (7) :

- C1.** Pour tout  $1 \leq i \leq n$ , on a  $\frac{f_{n_i}}{n_i} \rightarrow l \in \mathbb{R}$  lorsque  $n_i$  tends vers l'infini.

**C2.** L'application

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

est une fonction bijective.

**Théorème 2** *Sous les hypothèses (A1) à (A6), (C1) et (C2), on a, lorsque  $N$  tends vers l'infini, la normalité asymptotique suivante*

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

où  $\tilde{\boldsymbol{\theta}}^0 = (\tilde{\boldsymbol{\sigma}}^0, \boldsymbol{\beta}^0)$ , pour tout  $i \leq n$ ,  $\tilde{\sigma}_i^0 = \exp \left( \frac{f_{n_i}(\boldsymbol{\theta}^0)}{n_i} \right) \sigma_i^0$  et

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

Le Théorème 2 permet d'assurer la normalité asymptotique pour une toute nouvelle classe d'estimateurs  $\hat{\boldsymbol{\theta}}_f$ . Par conséquent, en utilisant une suite appropriée de fonctions  $(f_{n_i})_{1 \leq i \leq n}$ , nous pouvons modifier le problème d'optimisation (7) tout en conservant les mêmes garanties théoriques. Ici, nous nous intéressons au cas où la fonction  $f$  est une pénalisation de Ridge et nous voulons donc résoudre

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

Nous pouvons appliquer le Théorème 2 à (10) dans lequel  $f_{n_i}(\boldsymbol{\theta}) = \lambda \sigma_i^2$ , pour tout  $1 \leq i \leq n$  et  $\lambda > 0$  est un paramètre fixé. Nous avons la condition (C1) car, pour tout  $1 \leq i \leq n$ , nous avons  $\frac{f_{n_i}}{n_i} \longrightarrow 0$ . Dans ce cas, l'hypothèse (C2) est aussi vérifiée. En effet, si nous écrivons  $\tilde{\sigma}_i = \sigma_i \exp \left( \lambda \frac{\sigma_i^2}{n_i} \right)$ , nous pouvons obtenir les paramètres d'échelle désirés  $\sigma_i$  de la manière suivante :

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

avec  $W$  la fonction de Lambert, définie comme la fonction inverse de  $x \rightarrow xe^x$ .

En conséquence, l'estimateur défini par le problème d'optimisation (7) incluant un terme de pénalisation Ridge  $\lambda \sum_{i=1}^n \sigma_i^2$  est asymptotiquement normal, avec la même matrice de variance-covariance que le problème non pénalisé (car  $l = 0$  ici). L'ajout d'une telle pénalisation pourrait permettre un meilleur contrôle sur l'estimation des paramètres d'échelle du modèle. De plus, si  $Z \sim ALD(\mu, \sigma, \tau)$ , nous avons le résultat suivant (Yu and Zhang (2005))

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

Comme  $\sigma_i$  est le paramètre d'échelle pour le  $i$ -ième individu, on peut observer que la pénalisation de  $\sigma_i$  nous permet de contrecarrer la surestimation éventuelle de la variance entre les mesures d'un même individu. D'une certaine manière, l'information sur le paramètre d'échelle  $\sigma_i$  nous renseigne sur la dispersion ou la diversité des mesures pour un individu.

La méthode d'estimation de notre modèle n'est pas détaillé ici. Suivant Geraci (2014), notre approche s'appuie sur un algorithme d'optimisation itératif et l'utilisation d'une quadrature de Gauss. Une implémentation en R est disponible sur Github ici : <https://github.com/I621974/hlqmm>. La prochaine partie décrit quelques résultats de simulations justifiant des bonnes propriétés de nos estimateurs.

### 3 Étude par simulation

Cette section est consacrée à des expériences numériques afin d'évaluer la performance de notre méthode.

Pour chaque cas, nous considérons :  $p = 9$ ,  $J = 1000$ ,  $n = 6$ ,  $\tau = 0.5$ ,  $\psi = 2$ ,  $\boldsymbol{\sigma}^0 = (1, 2, 1, 3, 1, 0.5)^t$ ,  $\boldsymbol{\beta}^0 = (\psi, 1, 2, -2, -1, 3, 5.5, 0.5, 0.1)^t$ ,  $\phi = 5$  et pour tout  $i \in \{1, \dots, n\}$ ,  $n_i = J = 1000$ . Le paramètre  $\psi$  représente la valeur réelle de l'ordonnée à l'origine. Nous considérons ensuite trois distributions pour générer les covariables  $\mathbf{X}$ , à savoir une distribution normale  $\mathcal{N}(2, 2)$ , une distribution de Laplace  $\mathcal{L}(2, 3)$  et une distribution de Bernoulli  $\mathcal{B}(0.33)$ . Nous représentons l'histogramme pour  $\hat{\boldsymbol{\sigma}}$  et  $\hat{\boldsymbol{\beta}}$  de 1000 répliquions de Monte Carlo par coordonnées et on le compare à la distribution normale théorique asymptotique. Les Figures 1 et 2 montrent les distributions empiriques de  $\hat{\boldsymbol{\beta}}$  et  $\hat{\boldsymbol{\sigma}}$  respectivement lorsque les covariables sont générées selon la distribution normale  $\mathcal{N}(2, 2)$ .

Tout d'abord, nous pouvons constater que toutes les distributions empiriques correspondent aux distributions théoriques, à l'exception de l'ordonnée à l'origine. Il existe un biais sur l'ordonnée à l'origine, qui a été supprimé dans la Figure 1 en centrant l'estimateur de l'ordonnée à l'origine pour voir l'adéquation en termes de variance. On peut trouver quelques informations concernant la présence de ce biais et la nécessité d'inclure l'ordonnée à l'origine dans une régression quantile dans Jurečková (1984) et Battagliola et al. (2022). Sinon, nous pouvons voir que tous les estimateurs semblent suivent la distribution normale théorique ( $p > 0.05$ , test de Shapiro). Le comportement pour d'autres distributions de co-variables est en grande partie le même. Les figures relatives à ces distributions sont omises ici.

Nous nous intéressons ensuite à la convergence de notre estimateur  $\hat{\theta}$ . Nous utilisons le même *design* de simulation, mais nous faisons maintenant varier le nombre de mesures,  $J$ . Nous pouvons observer la convergence de l'algorithme dans la Figure 3 pour le paramètre  $\boldsymbol{\sigma}$  lorsque  $J$  augmente. Ici, les résultats sont assez similaires, quelle que soit la distribution utilisée pour générer les covariables. En regardant l'échelle sur l'axe des ordonnées, on constate que plus le paramètre est proche de 0, plus la méthode a du mal à l'estimer correctement. Les résultats sont les mêmes pour le paramètre  $\boldsymbol{\beta}$ .

$\mathcal{N}(2, 2)$

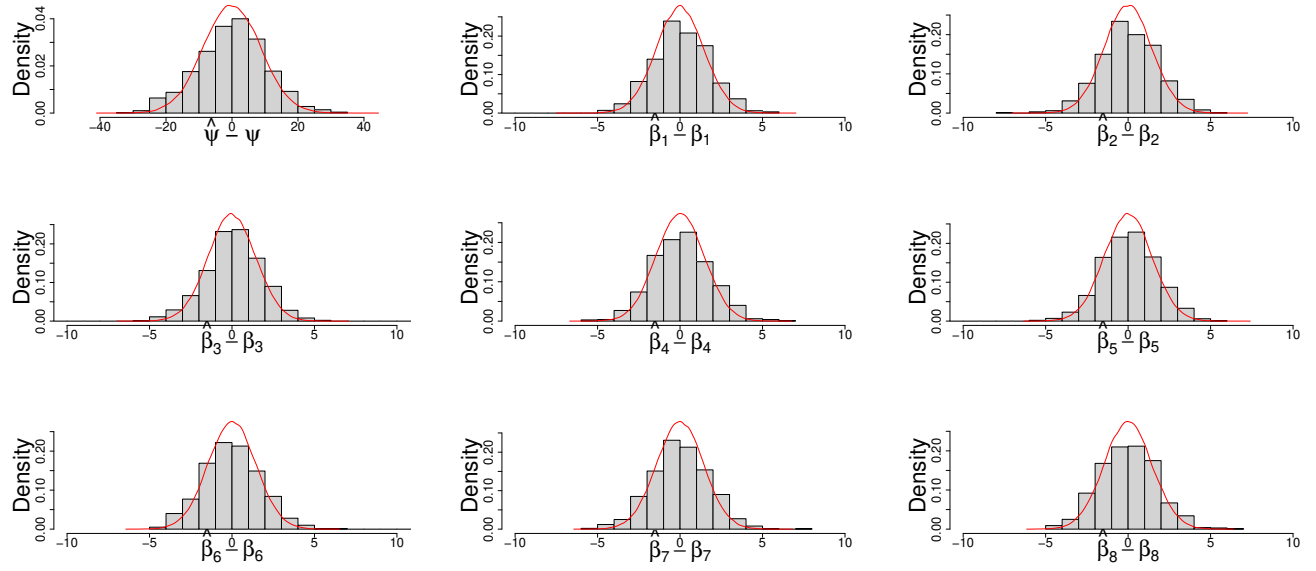


Figure 1: Distribution empirique de  $\hat{\beta} - \beta$  pour une génération normale de covariables suivant  $\mathcal{N}(2, 2)$ . Pour l'estimateur  $\hat{\psi}$ , le biais empirique a été supprimé.

$\mathcal{N}(2, 2)$

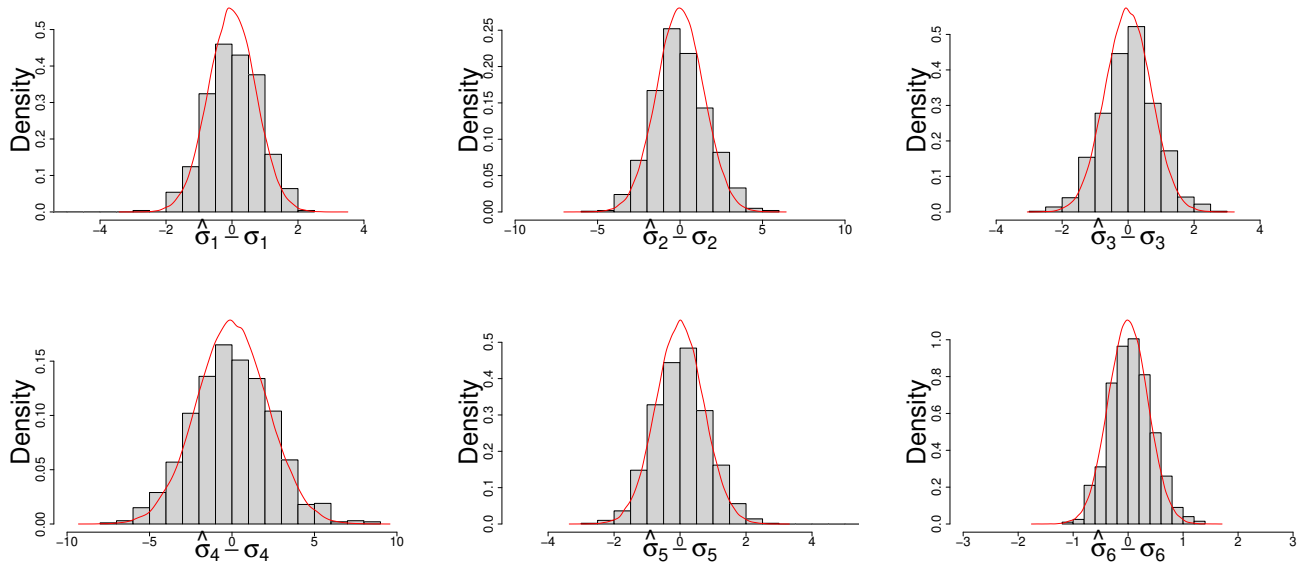


Figure 2: Distribution empirique de  $\hat{\sigma} - \sigma$  pour une génération normale de covariables suivant  $\mathcal{N}(2, 2)$ .

## 4 Conclusion

Cet article présente une nouvelle classe d'estimateurs basées sur la régression quantile pour modèle mixte qui permettent de préciser et maîtriser l'hétéroscédasticité dans des données



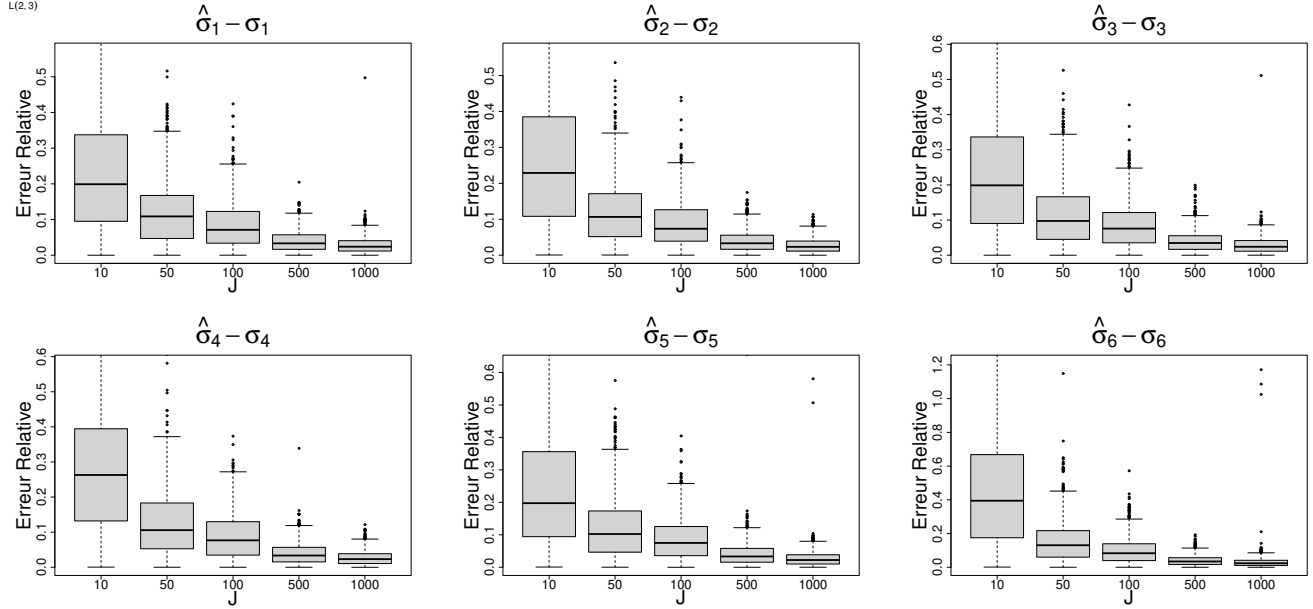


Figure 3: Illustration de la convergence de  $\hat{\sigma} - \sigma^0$  pour une génération de Laplace de co-variables suivant  $\mathcal{L}(2, 3)$ .

longitudinales. Les bonnes propriétés des estimateurs sont montrés théoriquement et soutenues par des simulations. L'intérêt de la méthode sera explicitée via une application écologique.

## Bibliographie

- Aghamohammadi, A. and Mohammadi, S. (2017), 'Bayesian analysis of penalized quantile regression for longitudinal data', *Statistical Papers* **58**, 1035–1053.
- 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.
- 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.
- 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.
- 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.
- 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.
- Jurečková, J. (1984), ‘Regression quantiles and trimmed least squares estimator under a general design’, *Kybernetika* **20**(5), 345–357.
- Koenker, R. (2004), ‘Quantile regression for longitudinal data’, *Journal of multivariate analysis* **91**(1), 74–89.
- Koenker, R. and Bassett, J. G. (1978), ‘Regression quantiles’, *Econometrica: journal of the Econometric Society* pp. 33–50.
- 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.
- 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.
- Weidenhammer, B. (2017), *The Consistency of Quantile Regression in Linear Mixed Models*, Freie Universitaet Berlin (Germany).
- 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.