Quantile and Expectile Regression for random effects model

Amadou Diogo Barry, Arthur Charpentier, Karim Oualkacha

Quantile and Expectile Regression for random effects model

Amadou Diogo Barry[1], Arthur Charpentier[2], and Karim Oualkacha[1]

[1]Department of Mathematics, Université du Québec à Montréal, Canada
[2]Falculty of Economics, Université de Rennes 1, France

December 20, 2016

Abstract

Quantile and expectile regression models pertain to the estimation of unknown quantiles/expectiles of the cumulative distribution function of a dependent variable as a function of a set of covariates and a vector of regression coefficients. Both approaches make no assumption on the shape of the distribution of the response variable, allowing for investigation of a comprehensive class of covariate effects. This paper fits both quantile and expectile regression models within a random effects framework for dependent/panel data. It provides asymptotic properties of the underlying model parameter estimators and suggests appropriate estimators of their variances-covariances matrices. The performance of the proposed estimators is evaluated through exhaustive simulation studies and the proposed methodology is illustrated using real data. The simulation results show that expectile regression is comparable to quantile regression, easily computable and has relevant statistical properties. In conclusion, expectiles are to the mean what quantiles are to the median, and they should be used and interpreted as quantilized mean.

Keywords: Expectiles, quantiles, random effects, longitudinal, panel data.

1 Introduction

With development and progress of computer science, quantile regression (QR) and expectile regression (ER) extended the scope of statistical modeling beyond the classical linear regression of the conditional mean. QR extends the usual univariate quantile to the conditional quantile class, by estimating the conditional quantile of the dependant variable. Similarly, ER extends the univariate expectile to the conditional expectile class. Conditional quantile and conditional expectile retain all the properties inherent to their univariate counterpart. QR and ER are suitable in situations where the effect of an explanatory factor on the conditional mean or median does not capture the impact of the factor in the whole distribution of the response variable. In such situations, the factor does not affect all quantiles/expectiles of the response variable in the same way. QR and ER have similar roles in modeling, as both provide a thorough and detailed insight of the influence of risk factors on the distribution of the dependent variable. However, they are distinguished by their properties, advantages and disadvantages.

Quantile of level α of a variable is a common descriptive statistic. Quartiles are among the most familiar and most common quantiles. In a descriptive analysis, they provide a complete picture of the distribution of the variable. Until recently, the advantage of laying out a complete picture of the distribution of a variable from a few statistics was only accessible in the univariate case. The relationship of multiple variables was studied by the estimation of the conditional mean. Today, with the introduction of the QR, it is possible to study the impact of a factor, taking into account other factors, not only on the conditional mean, but also upon other functions of the distribution of the dependent variable.

QR estimator or weighted asymmetric least absolute deviation estimator was introduced in 1978 by Koenker and Basset [20] to analyse the relationship between the conditional quantiles of the response distribution and a set of regressors. Since its appearance, its theoretical and empirical development have continuously increased. Today, its scope span all areas of applied science [34]. A large part of QR literature focuses on solving problems related to estimation of the variances-covariances matrix of the QR estimators and their performance on small samples [2]. There are also many studies that have adapted the QR to different models used previously for the estimation of the conditional mean. Powell [26] adapted the QR to censored data and Machado [23] generalized it to count data. Koenker and Bilias[14], Fitzenberger and Wilke [6] applied QR to duration data. Lately, QR was adjusted

to panel and longitudinal data [16, 3, 22]. For further details on the theory and application of the QR, see [17].

Weighted asymmetric least square deviation estimator (expectile), was initially studied in [1]. However, it is in [25] that it is named "expectile" for the first time, although other qualifications (gravile, heftile, loadile, projectile) were used elsewhere in the past to identify it. The best well known expectile is the expectile of level $\tau = 0.5$ corresponding to the mean. Aside from the mean, the other expectiles are less well known, and are not as easily interpretable as the mean or the quantiles. However, they are a reliable alternative to the quantiles. Similarly to quantiles, a sequence of expectiles can be sufficient to describe the distribution of a variable, especially with the mean and a few expectiles above and below the mean.

The usefulness of this new class of estimator and its resemblance to the class of weighted asymmetric least absolute deviation (or quantile) estimator were put forward in [5]. Today, there is growing interest in the literature [13, 33, 31, 29] toward the expectile estimator. This increased interest is explained partly by the close link between expectiles and quantiles, and also by the attractive properties of expectiles and the limits of quantiles. Unlike QR estimators, ER estimators have explicit form and are analytically estimable. Moreover, the overlap problem of QR estimator is less common with ER estimator [33].

Panel data also known as longitudinal data or dependant data is by far the most appreciated observational data. Longitudinal data are characterized by recording measurements of the same individuals repeatedly through time. They offer the opportunity to control unobserved individual heterogeneity. Longitudinal data arises in many application studies such as in econometrics [12], epidemiology [28], genetic [7], etc.

Random effect and fixed effect models are the most popular approaches used in econometrics to adjust panel data [12]. QR model has been generalized to several models for panel or longitudinal data, fixed effects model [3], instrumental variable model [11], linear mixed model [10], penalized fixed effect model [16, 22], among others. Despite the ubiquity of the classical conditional mean regression, QR has become a standard model. In the meantime generalization of the ER is not effective, regardless of its desirable properties. The aim of this paper is to adapt QR and ER to random effects model for panel data. It provides asymptotic properties of the underlying model parameter estimators and suggests appropriate estimators of their variances-

covariances matrices. To the best of our knowledge it is the first time that ER estimator and its asymptotic properties are considered for random effects model.

The paper is organized as follows. Section 2 introduces the univariate quantile and expectile functions. Section 3 presents the asymptotic properties of the QR and the ER of the linear random effect model with suggested estimators of their variances-covariances matrix. Section 4 presents the performance of the proposed methodology in practice through simulations and a real data analysis. Section 5 contains the conclusion and Section 6 provides proofs of the theorems.

2 Quantiles and Expectiles

2.1 Quantiles

The quantile of level $\alpha \in [0,1]$ of a random variable Y is defined by

$$q(\alpha, Y) = F_Y^{-1}(\alpha) = \inf\{y; F_Y(y) \geq \alpha\},$$

where F_Y is the cumulative distribution function (c.d.f.) of Y. The quantile function $q(\cdot, Y)$ characterises the c.d.f. F_Y. For intance, when $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $q(\alpha, Y) = \mu + \sigma\Phi^{-1}(\alpha)$, where Φ denotes the c.d.f. of $\mathcal{N}(0,1)$.

It is known that the αth quantile can be also defined as the minimizer of the following expected loss

$$q(\alpha, Y) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\{r_\alpha^Q(Y - \theta)\},$$

where $r_\alpha^Q(u) = |\alpha - \mathbb{1}(u \leq 0)| \cdot |u|$ is the so-called check function. Notice that $q(\alpha, Y)$ is unique when the c.d.f F_Y is absolutely continuous.

Given a random sample $\{y_1, \cdots, y_n\}$, the αth sample quantile estimate can be obtained by sorting and ordering the n observations. It can be also obtained as

$$\begin{aligned}\widehat{q}(\alpha, \boldsymbol{y}) &= \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \int r_\alpha^Q(y - \theta) dF_n(y) \\ &= \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} r_\alpha^Q(y_i - \theta) \right\},\end{aligned}$$

with $F_n(.)$ stands for the empirical distribution function of Y and $\boldsymbol{y} = (y_1, \cdots, y_n)^\mathsf{T}$ is the $n \times 1$ random sample vector.

Similarly, conditional αth quantile of $Y|\boldsymbol{x}$ can be defined as

$$q(\alpha, Y, \boldsymbol{x}) = F^{-1}_{Y|\boldsymbol{x}}(\alpha) = \inf\{y; F_{Y|\boldsymbol{x}}(y) \geq \alpha\},$$

where $\boldsymbol{x}$ is a $p \times 1$ random (explanatory) vector and $F_{Y|\boldsymbol{x}}$ is the conditional c.d.f. of $Y|\boldsymbol{x}$. If one assume that the conditional quantile is a linear function of $\boldsymbol{x}$ (i.e. $F^{-1}_{Y|\boldsymbol{x}}(\alpha) = \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}^Q(\alpha)$), then the sample conditional αth quantile estimate can be obtained by solving the following optimization problem, with respect to $\boldsymbol{\beta}^Q(\alpha)$

$$\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n} \sum_{i=1}^{n} r^Q_\alpha(y_i - \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}^Q(\alpha)) \right\}, \tag{1}$$

with $\{(y_1, \boldsymbol{x}_1), \cdots, (y_n, \boldsymbol{x}_n)\}$ is a multivariate random sample and $\boldsymbol{X} = [\boldsymbol{x}_1; \cdots; \boldsymbol{x}_n]^\mathsf{T}$ is the $n \times p$ design matrix. This is well known as the quantile regression model [17].

Notice that $\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X})$ the solution of (1) can be also viewed as maximum likelihood estimator of $\boldsymbol{\beta}^Q(\alpha)$ when the disturbances, $\epsilon_i(\alpha) = y_i - \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}^Q(\alpha)$, arise from the asymmetric Laplace distribution [9, 10].

Quantiles have attractive equivariance properties that are very useful in reducing the computation time of the algorithms. For example, $q(\alpha, h(Y)) = h(q(\alpha, Y))$ if h is an increasing function. In particular, $q(\alpha, sY + t) = sq(\alpha, Y) + t$, with $(s, t) \in \mathbb{R}^+ \times \mathbb{R}$. With the multivariate random sample we have the following properties [20]:

- $\widehat{\boldsymbol{\beta}}^Q(\alpha, \lambda\boldsymbol{y}, \boldsymbol{X}) = \lambda\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}), \quad \lambda \in [0, \infty)$,
- $\widehat{\boldsymbol{\beta}}^Q(1-\alpha, \lambda\boldsymbol{y}, \boldsymbol{X}) = \lambda\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}), \quad \lambda \in (-\infty, 0]$,
- $\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y} + \boldsymbol{X}\gamma, \boldsymbol{X}) = \widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}) + \gamma, \quad \gamma \in \mathbb{R}^p$,
- $\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}A) = A^{-1}\widehat{\boldsymbol{\beta}}^Q(\alpha, \boldsymbol{y}, \boldsymbol{X}), \quad A_{p\times p}$ is nonsingular.

2.2 Expectiles

The expectile of level $\tau \in [0, 1]$ of a random variable Y is defined by

$$\mu(\tau, Y) = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\{r^E_\tau(Y - \theta)\} \text{ with } r^E_\tau(u) = |\tau - \mathbb{1}(u \leq 0)| \cdot u^2. \tag{2}$$

Similarly to the quantile function $q(\cdot, Y)$, the expectile function $\mu(\cdot, Y)$ characterises the c.d.f of Y. The following equation summarizes such a relationship

$$\mu(\tau, Y) = \mu - \frac{1-2\tau}{1-\tau}\,\mathbb{E}\left[\{Y - \mu(\tau, Y)\}\mathbb{1}\{Y > \mu(\tau, Y)\}\right], \tag{3}$$

with $\mu = \mu(0.5, Y) = \mathbb{E}(Y)$. For example, if $Y \sim \mathcal{N}(\mu, \sigma^2)$ then $\mu(\tau, Y) = \mu(\tau)$ is the solution of the following equation:

$$(2\tau-1)\mu(\tau)\Phi\left(\frac{\mu(\tau)-\mu}{\sigma}\right)+\mu = \tau\mu(\tau)+(1-\tau)\lambda\left(-\frac{\mu(\tau)-\mu}{\sigma}\right)-\tau\lambda\left(\frac{\mu(\tau)-\mu}{\sigma}\right)$$

where $\lambda(x) = \phi(x)/(1-\Phi(x))$ denotes the hazard function and $\phi(x)$ the standard normal density function. Expectile is location and scale equivariant. In fact, for $s > 0$ and $t \in \mathbb{R}$, one has

$$\mu(\tau, sY + t) = s\mu(\tau, Y) + t,\ s > 0 \text{ and } t \in \mathbb{R}.$$

An empirical estimate of (2) can be derived as a solution of

$$\widehat{\mu}(\tau, \boldsymbol{y}) = \underset{\theta\,\in\,\mathbb{R}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} r_\tau^E(y_i - \theta) \right\}.$$

Conditional τth expectile of Y given $\boldsymbol{x}$ can be defined as

$$\mu(\tau, Y, \boldsymbol{x}) = \underset{\theta\,\in\,\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\{r_\tau^E(Y-\theta)|\boldsymbol{x}\}.$$

If one assume that $\mu(\tau, \boldsymbol{x}) = \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta}^E(\tau)$ is a linear function in $\boldsymbol{x}$, then the derivation of a sample conditional τth expectile estimate leads to the following sample estimate of $\boldsymbol{\beta}^E(\tau)$

$$\widehat{\boldsymbol{\beta}}^E(\tau, \boldsymbol{y}, \boldsymbol{X}) = \underset{\boldsymbol{\beta}\,\in\,\mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n} r_\tau^E(y_i - \boldsymbol{x}_i{}^\mathsf{T}\boldsymbol{\beta}^E(\tau)) \right\}.$$

This minimization problem with respect to $\boldsymbol{\beta}^E(\tau)$ is known as expectile regression. One note that this estimator can also be derived by a likelihood-based approach from a Gaussian density with unequal weights placed on positive and negative disturbances [1].

Note that several characteristics and expressions of quantiles have there analogous for expectiles. For instance, for quantiles we have

$$\alpha = F(q(\alpha)) = \mathbb{E}[\mathbb{1}(Y < q(\alpha))]. \tag{4}$$

Re-arrangement of equation (3) leads to analogous expression for expectiles

$$\tau = \frac{\mathbb{E}[|Y - \mu(\tau)|\mathbb{1}\{Y < \mu(\tau)\}]}{\mathbb{E}[|Y - \mu(\tau)|]}. \tag{5}$$

Equations (4) and (5) show that expectiles are determined by tail expectations of Y, while quantiles are determined by the distribution function.

Moreover, one can show that for each αth quantile $q(\alpha)$ there is a unique $\tau_\alpha th$ expectile, $\mu(\tau_\alpha)$, such that $\mu(\tau_\alpha) = q(\alpha)$. One can write then

$$\tau_\alpha = \frac{q(\alpha) - \mathbb{E}[Y\mathbb{1}\{Y < q(\alpha)\}]}{\mu - 2\,\mathbb{E}[Y\mathbb{1}(Y < q(\alpha))] - q(\alpha)(1 - 2\alpha)}.$$

Furthermore, when the c.d.f. function of Y is given by

$$F(y) = 1/2\{1 + \text{sign}(y)\sqrt{1 + 4/(4 + y^2)}\}, \quad y \in \mathbb{R}$$

the αth quantile coincides with the αth expectile. That is, $\tau_\alpha = \alpha$, [15]. Finally, we close this section with a comparison between quantile and expectile functions of the standard normal distribution, given in Figure 1. The two functions intersect at point $\alpha = \tau = 0.5$. Both functions are strictly increasing and cover all values of $I_F = \{y|0 < F(y) < 1\}$.

Given all these analogous relationships between quantiles and expectiles and the fact that the mean is a particular expectile with $\tau = 0.5$, one can interpret expectiles as quantilized means.

3 Quantile and Expectile regression for panel data

Panel data is by far the most appreciated observational data. They offer the opportunity to control unobserved individual heterogeneity. Panel data are characterized by recording multiple observations on individuals over several periods.

A standard panel-data linear regression model relies a response y to predictors as follow

$$y_{it} = \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta} + u_i + v_{it}, \qquad t \in \{1, 2 \ldots, T_i\}, \qquad i \in \{1, 2 \ldots, n\} \tag{6}$$

with y_{it} is a dependent variable, $\boldsymbol{x}_{it} = (x_{it}^1, x_{it}^2, \ldots, x_{it}^p)^{\mathsf{T}}$ is a vector of p independent variables measured on individual i at time t, u_i an unobserved

individual-specific effect and v_{it} an error term.

Fixed effects and random effects methods are most popular approaches to adjust model (6) in econometrics. Both approaches have advantages and limitations and the comparison between the two is delicate and has been widely discussed in the literature [12].

3.1 Quantile regression for fixed effects model

In the fixed effect (FE) paradigm the effects of unobserved heterogeneity are assumed to be fixed parameters and the model (6) is conveniently written as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{v}. \tag{7}$$

where the vector $\boldsymbol{y}$ is $N \times 1$, $\boldsymbol{X}$ is $N \times p$ matrix, $\boldsymbol{Z}$ is a $N \times n$ incidence matrix and $N = \sum_{i=1}^{n} T_i$. The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are respectively $N \times 1$ individual-specific effect and error vectors. The corresponding QR fixed effects (QRFE) model proposed by Koenker [16] is formulated as

$$q(\alpha, y_{it}, \boldsymbol{x}_{it}) = u_i + \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\alpha). \tag{8}$$

The individual-specific effects, u_i, are treated as fixed parameters in the model and have to be estimated in addition to the structural parameters $\boldsymbol{\beta}(\alpha)$. The QRFE model parameters can be obtained by solving

$$\min_{\boldsymbol{\beta},\boldsymbol{u}} \sum_{i=1}^{n} \sum_{t=1}^{T_i} r_{\alpha}^{Q}\{y_{it} - u_i - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\alpha)\}.$$

Because of the dimension of the incidental parameters, u_i, estimation is challenging particularly as $n \to \infty$. Also, adaptation of techniques used for the FE approach from classical least square estimation are not directly applicable in the QR context. To remedy this problem, Koenker [16] exploited the sparse property of the full-model design-matrix $(\boldsymbol{X}, \boldsymbol{Z})$ to reduce the computational burden problem of QRFE model. He also derived asymptotic properties of the model parameter estimators as n and $T_i \to \infty$.
Koenker introduced a more general class for the QRFE model, namely a class of penalized QRFE model by solving

$$\min_{\boldsymbol{\beta},\boldsymbol{u}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T_i} w_k r_{\alpha_k}^{Q}\{y_{it} - u_i - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\alpha_k)\} + \lambda \sum_{i=1}^{n} |u_i|,$$

where w_k is a relative weight associated to the q quantiles $\{\alpha_1, \ldots \alpha_q\}$ and λ is a shrinkage parameter. Koenker used the l_1 penalty for its advantages

over the l_2 penalty. In fact, the l_1 penalty serves to shrink the incidence parameter toward zero and then provides sparse solutions for estimates of u_i, and at the same time improves the performance of the estimate of the structural parameters.

The choice of the shrinkage parameter, λ, influences heavily the inference for the parameter of interest [9, 22]. For example, the case $\lambda \to \infty$ is equivalent to the ordinary QR, while $\lambda \to 0$ coresponds to the QRFE model. Lamarche [22] focused on the problem of selecting the optimal regularization parameter, and Canay [3] suggested a consistent and asymptotically normal two-stage estimator for $\boldsymbol{\beta}(\alpha)$, which copes with the incidental parameters curse in the first stage.

3.2 Quantile regression for random effects model

In this section we propose to handle panel data using a QR random effect (QRRE) model. Our main contribution in this section is the derivation of asymptotic distribution of the QRRE model parameter estimators and provide a consistent estimator of their variances-covariances matrix.

The random effect model or the error component model is a particular type of random coefficients model also known as multilevel model, hierarchical linear model or linear mixed model [8]. In the random effect framework, unobserved heterogeneity effects are treated as random variables and are assumed to be uncorrelated to the explanatory variables. Thus, the linear panel equation (6) can be reformulated in matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{9}$$

where the random vector $\boldsymbol{\varepsilon}$ is the sum of the individual-specific effects and the disturbance vector, $\boldsymbol{\varepsilon} = \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{v}$. The corresponding conditional quantile regression model can be written as

$$q(\alpha, y_{it}, \boldsymbol{x}_{it}) = \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}^{Q}(\alpha), \qquad t \in \{1, 2 \ldots, T_i\}, \qquad i \in \{1, 2 \ldots, n\}.$$

The parameter $\boldsymbol{\beta}^{Q}(\alpha)$ models the effect of the independent variables on the location, scale and shape of the conditional distribution of the response. For example, with one regressor and under the absence of individual-specific effects (i.e. under homoskedasticity condition with ε_i are independent and identically distributed, i.i.d.), quantile functions $q(\alpha, y_{it}, \boldsymbol{x}_{it}) = \beta_0^{Q}(\alpha) + x_{it}\beta_1^{Q}$ are parallel lines with $\beta_0^{Q}(\alpha) = \beta_0^{Q} + F_{\varepsilon}^{-1}(\alpha)$, with $F_{\varepsilon}(.)$ is the c.d.f function of the error terms. In the presence of individual-specific effects, $\beta_1^{Q}(\alpha)$ will

vary with α.

The QRRE model parameters can be estimated by solving

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T_i} r_\alpha^Q (y_{it} - \boldsymbol{x}_{it}^\mathsf{T} \boldsymbol{\beta}^Q(\alpha)). \tag{10}$$

This minimization problem is known as weighted asymmetric least absolute deviation (WALAD) estimator of $\boldsymbol{\beta}^Q(\alpha)$ and the solution can be obtained using existing R packages such as [18]. In the next section we will derive the asymptotic properties of the WALAD estimator of $\boldsymbol{\beta}^Q(\alpha)$.

From now one we will restrict attention to balanced data ($T_i = T$). We will assume i.i.d. individual vectors $\{\boldsymbol{y}_i,\ i \in (1, \cdots, n)\}$, and also assume independence of the individual-specific effects u_i and the disturbances v_{it}.

3.2.1 Asymptotic properties of QRRE model

Assume the following regularity conditions:

Condition Q1. Y_{it} is a continuous random variable with absolutely continuous distribution function F_{it} and with a continuous density function f_{it} uniformly bounded away from 0 and ∞ at the points $q(\alpha, y_{it})$.

Condition Q2. There are positive definite matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ such that:

1. $\lim_{n\to\infty} N^{-1} \boldsymbol{X}^\mathsf{T} \{\mathbb{I}_n \otimes \Sigma_{T\times T}(\alpha)\} \boldsymbol{X} = \mathbf{D}_0$

2. $\lim_{n\to\infty} N^{-1} \boldsymbol{X}^\mathsf{T} \Omega_f \boldsymbol{X} = \mathbf{D}_1$

3. $\max_{1\le i\le n, 1\le t\le T} \|\boldsymbol{x}_{it}\|_2 / \sqrt{N} \to 0$

and for every $i \in \{1, \ldots, n\}$ the elements of $\Sigma_{T\times T}(\alpha)$ are defined by

$$\sigma_{its}(\alpha) = \begin{cases} \alpha(1-\alpha) & \text{if } t = s \\ \mathbb{E}[\mathbb{1}\{\varepsilon_{it}(\alpha) < 0, \varepsilon_{is}(\alpha) < 0\}] - \alpha^2 & \text{if } t \neq s \end{cases}$$

and

$$\Omega_f = \text{Diag}\big[f_{11}\{q(\alpha, y_{11})|\boldsymbol{x}_1\}, \ldots, f_{1T}\{q(\alpha, y_{1T})|\boldsymbol{x}_1\}, \ldots, f_{n1}\{q(\alpha, y_{n1})|\boldsymbol{x}_n\}, \ldots, f_{nT}\{q(\alpha, y_{nT})|\boldsymbol{x}_n)\}\big].$$

Our assumptions are standard in the literature of the QR model. The only new assumption is related to the individual-random effect which introduces dependency between observations of the same individual. Under the above set of conditions, we present the main results of the QRRE model parameter estimators.

Theorem 3.1. *Assume model (9) satisfies conditions **Q1** and **Q2**. Let $\widehat{\boldsymbol{\beta}}^Q(\alpha)$ be an estimate of the true parameter $\boldsymbol{\beta}^Q(\alpha)$ obtained from the minimization of the loss function in (10). The following result holds*

$$\sqrt{N}\Big(\widehat{\boldsymbol{\beta}}^Q(\alpha) - \boldsymbol{\beta}^Q(\alpha)\Big) \xrightarrow{d} \mathcal{N}\Big(0, \mathbf{D}_1^{-1}\mathbf{D}_0\mathbf{D}_1^{-1}\Big).$$

Note that the variances-covariances matrix of the WALAD estimator depends on the density f_{it} and on the joint distribution function F_{its} of y_{it} and y_{is}, for $i \in \{1, \ldots, n\}$ and $t, s \in \{1, \ldots, T\}$. Hence, estimating standard errors, variances-covariances matrix as well as confidence intervals of the WALAD estimator require estimating $f_{it}\{q(\alpha, y_{it})\}$, and $F_{its}\{q(\alpha, y_{it}), q(\alpha, y_{is})\}$. Estimation of the density function has been extensively studied [24] in the context of constructing confidence interval and test for inference of linear quantile regression estimator.

Now we present an algorithm for estimating the variances-covariances matrix of the WALAD estimator.

1. Estimate $\widehat{\varepsilon}_{it}(\alpha) = y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}^Q(\alpha)$ and $\widehat{q}(\alpha, y_{it}) = \boldsymbol{x}_{it}{}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}^Q(\alpha)$, with $\widehat{\boldsymbol{\beta}}^Q(\alpha)$ the estimator of Theorem 3.1.

2. Estimate $\widehat{f_{it}}$ by one of the consistent estimator of f_{it} suggested in the literature [24].

3. Estimate the matrix elements $\sigma_{its}(\alpha)$ of $\Sigma_{T\times T}(\alpha)$ by

$$\widehat{\sigma}_{its}(\alpha) = \begin{cases} \alpha(1-\alpha) & \text{if } t = s \\ \frac{1}{T(T-1)(n-1)} \sum_{i=1}^{n} \sum_{l<m}^{T} \mathbb{1}(\widehat{\varepsilon}_{il}(\alpha) < 0, \widehat{\varepsilon}_{im}(\alpha) < 0) - \alpha^2 & \text{if } t \neq s. \end{cases}$$

4. Estimate $\mathbf{D}_0$ by

$$\widehat{\mathbf{D}}_0 = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \boldsymbol{x}_{it}\widehat{\sigma}_{its}(\alpha)\boldsymbol{x}_{is}{}^{\mathsf{T}}.$$

5. And estimate $\mathbf{D}_1$ by

$$\widehat{\mathbf{D}}_1 = N^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \widehat{f}_{it}(\widehat{q}(\alpha, y_{it})|\boldsymbol{x}_i)\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}.$$

With the estimators $\widehat{\mathbf{D}}_0$ and $\widehat{\mathbf{D}}_1$ of the algorithm, we have the following Theorem:

Theorem 3.2. *Under Assumptions **Q1** and **Q2**, for every $\alpha \in (0,1)$ we have*

$$\widehat{\mathbf{D}}_1^{-1}\widehat{\mathbf{D}}_0\widehat{\mathbf{D}}_1^{-1} \xrightarrow{\mathbb{P}} \mathbf{D}_1^{-1}\mathbf{D}_0\mathbf{D}_1^{-1}.$$

The proofs of Theorems 3.1 and 3.2 are postponed in Section 5.

3.3 Expectile regression for random effects model (ERRE)

In this section we derive the asymptotic properties of the ERRE model parameter estimators and suggest a consistent estimator of their variances-covariances matrix. The asymptotic properties for the classical expectile linear regression have been proven by Newey and Powell [25] and Sobotka et al. [30] has provided the results for the semiparametric regression model estimators.

Let the conditional expectile regression of the random effect model (9) defined as:

$$\mu(\tau, y_{it}, \boldsymbol{x}_{it}) = \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}^E(\tau), \qquad t \in \{1, 2 \ldots, T\}, \qquad i \in \{1, 2 \ldots, n\},$$

for every $\tau \in (0,1)$. As with the quantile regression, under a linear homoskedastic regression with one regressor the expectile functions $\mu(\tau, y_{it}, x_{it}) = \beta_0^E(\tau) + x_{it}^{\mathsf{T}}\beta_1^E$ are parallel lines with $\beta_0^E(\tau) = \beta_0^E + \mu(\tau, \varepsilon)$.

The expectile regression estimator of the random effects model (9) is defined as solution of:

$$\underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{n} \sum_{t=1}^{T} r_\tau^E\big\{y_{it} - \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}^E(\tau)\big\}. \tag{11}$$

Contrary to the quantile regression estimator, the expectile regression estimator has an explicit form, computable as iterated weighted least squares estimators, given by:

$$\widehat{\boldsymbol{\beta}}^{E}(\tau) = \Big(\sum_{i=1}^{n}\sum_{t=1}^{T} \widehat{w}_{i,t}(\tau)\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}\Big)^{-1}\Big(\sum_{i=1}^{n}\sum_{t=1}^{T} \widehat{w}_{i,t}(\tau)\boldsymbol{x}_{it}y_{it}\Big),$$

with $\widehat{w}_{it}(\tau) = |\tau - \mathbb{1}(y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}^{E}(\tau))|$. Since the weighted asymmetric quadratic loss is convex and differentiable, traditional procedure can be applied to derive the asymptotic properties of the ERRE model parameter estimators.

3.3.1 Asymptotic properties of ERRE model

The asymptotic properties of this new estimator is presented under the assumptions state below. Without loss of generality, let d be a generic constant.

Assumption E1. For each sample $\boldsymbol{z}_{it} = (y_{it}, \boldsymbol{x}_{it}^{\mathsf{T}})$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$ of size $N = nT$, with T fixed, we assume $\boldsymbol{y}_i = (y_{it}, \ldots, y_{iT})^{\mathsf{T}}$ is i.i.d and has a continuous probability density function $f(\boldsymbol{y}_i|\boldsymbol{x}_i)$.

Assumption E2. There is a constant $d > 0$ and a measurable function $\alpha(\boldsymbol{z})$ such that: $f(\boldsymbol{y_i}|\boldsymbol{x}_i) < \alpha(\boldsymbol{z})$ and:

$$\int \|\boldsymbol{z}\|^{4+d}\alpha(\boldsymbol{z})d\mu_{\boldsymbol{z}} < +\infty, \quad \int \alpha(\boldsymbol{z})d\mu_{\boldsymbol{z}} < +\infty.$$

Assumption E3. $\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}$ is nonsingular.

Our assumptions are relatively identical to those set out in [25]. The main difference is the exclusion of a misspecification parameter denoted γ in [25] and the related assumptions.

For $\varepsilon_{it}(\tau) = y_{it} - \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau)^{E}$ and $w_{it}(\tau) = |\tau - \mathbb{1}(\varepsilon_{it}(\tau) < 0)|$, let

$$\begin{aligned}
\overline{W} &= \mathbb{E}(W), \quad W = \mathrm{Diag}(w_{11}(\tau), \ldots w_N(\tau)) \\
H &= \boldsymbol{X}^{\mathsf{T}}\overline{W}\boldsymbol{X} \\
\Sigma &= \boldsymbol{X}^{\mathsf{T}}\,\mathbb{E}(W\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathsf{T}}W)\boldsymbol{X}.
\end{aligned}$$

Under the above set of conditions, we present the main results of the ERRE model parameter estimators.

Theorem 3.3. *Assume model (9) satisfies hypotheses **E1-E3**. Let $\widehat{\boldsymbol{\beta}}^E(\tau)$ be the estimator of the true parameter $\boldsymbol{\beta}^E(\tau)$ obtained by minimizing the loss function*

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T} r_\tau^E\{y_{it}-\boldsymbol{x}_{it}{}^\intercal\boldsymbol{\beta}(\tau)^E\}$$

then, for every $\tau\in(0,1)$, we have

$$\sqrt{N}\{\widehat{\boldsymbol{\beta}}^E(\tau)-\boldsymbol{\beta}^E(\tau)\}\xrightarrow{d}\mathcal{N}(0,H^{-1}\Sigma H^{-1}).$$

As for the QRRE estimator we present an algorithm for estimating the variances-covariances matrix of the ERRE estimator.

1. Estimate $\widehat{\varepsilon}_{it}(\tau)=y_{it}-\boldsymbol{x}_{it}{}^\intercal\widehat{\boldsymbol{\beta}}^E(\tau)$ and $\widehat{w}_{it}(\tau)=|\tau-\mathbb{1}(\widehat{\varepsilon}_{it}(\tau)<0)|$ with $\widehat{\boldsymbol{\beta}}^E(\tau)$ the estimator of Theorem 3.3.
2. Compute $\widehat{W}=\mathrm{Diag}(\widehat{w}_{11}(\tau),\ldots,\widehat{w}_N(\tau))$.
3. Compute $\widehat{H}=\boldsymbol{X}^\intercal\widehat{W}\boldsymbol{X}/N$.
4. Compute $\widehat{\Sigma}=\boldsymbol{X}^\intercal\widehat{W}\widehat{\boldsymbol{\varepsilon}}\widehat{\boldsymbol{\varepsilon}}^\intercal\widehat{W}\boldsymbol{X}/N$.

With the estimators $\widehat{W}$, $\widehat{H}$ and $\widehat{\Sigma}$, we have the following Theorem:

Theorem 3.4. *Under Assumptions **E1-E3**, for every $\tau\in(0,1)$ we have*

$$\widehat{H}^{-1}\widehat{\Sigma}\widehat{H}^{-1}\xrightarrow{\mathbb{P}}H^{-1}\Sigma H^{-1}.$$

The proofs of Theorems 3.3 and 3.4 are presented in Section 5.

4 Simulation and Application

In the previous section we presented the asymptotic properties of QRRE and ERRE estimators for the random effects model. In this section, we study their performance in practice through simulations and real data. Performance comparisons of the proposed approaches are realized with respect to the penalized quantile regression (PQR) fixed effect approach [16].
From now on we replace α by τ and write $q(\tau)$ for quantiles.

4.1 Simulation

4.1.1 Design

Our simulation study follows closely the different scenarios proposed in [16]. The simulated data is generated under two different models: a location shift and a location-scale shift models given by

$$y_{it} = \begin{cases} x_{it}\beta + u_i + v_{it}, & \text{location shift.} \\ x_{it}\beta + u_i + (1 + x_{it}\gamma)v_{it}, & \text{location-scale shift.} \end{cases}$$

Regardless which model are used to generate the data, the corresponding QRRE τth quantile and ERRE τth expectile are respectively:

$$\begin{cases} q(\tau, y_{it}) = & \beta_0^Q(\tau) + x_{it}\beta_1^Q(\tau), \\ \mu(\tau, y_{it}) = & \beta_0^E(\tau) + x_{it}\beta_1^E(\tau). \end{cases}$$

The individual random effect u_i and the disturbance v_{it} are generated by the same distribution in three different models: normal distribution $\mathcal{N}(0,1)$, Student distribution t_3 with 3 degree of freedom, and central chi-squared distribution χ_3^2 with 3 degree of freedom. The continuous explicative variable x is generated by a normal distribution $\mathcal{N}(0,1)$ in the location shift model and by a central chi-squared distribution χ_3^2 with 3 degree of freedom in the location-scale shift model. We set $\beta = 0$ and $\gamma = 1/10$.

According to the different models and the different sample sizes $n \times T \in \{50, 100, 500\} \times \{5, 15, 15\}$, we have created 54 different random samples. The three methods ERRE, QRRE and PQR, were fitted to each sample replication to estimate quantiles and expectiles of levels $\tau \in \{0.1, 0.2, 0.5, 0.8, 0.9\}$. The PQR approach requires the tuning parameter λ. We set the tuning parameter to be the ratio of scale parameters, i.e. $\lambda = \sigma_v/\sigma_u$, as suggested by [16]. With $\sigma_u^2 = \mathrm{Var}(u_i)$ and $\sigma_v^2 = \mathrm{Var}(v_{it})$ in the location shift model and $(1 + x_{it}\gamma)^2 \mathrm{Var}(v_{it})$ in the location scale-shift scenario.

For the measurement of the quality of the different methods we calculated the bias $\left(1/m \sum^m \widehat{\beta}^{(r)}\right)$ and the root-mean-square error $\left(\sqrt{1/m \sum^m (\widehat{\beta}^{(r)} - \bar{\beta})^2}\right)$ where $\bar{\beta} = 1/m \sum^m \widehat{\beta}^{(r)}$. The reported results are based on $m = 400$ replications.

As we simulated several samples (54), we estimated a small series of asymmetric point $\Big(\tau \in \{0.1, 0.2, 0.5, 0.8, 0.9\}\Big)$, and reserved the longer series of

asymmetric points to the real data set. All results are presented in Figure **??** and Figure 3, while only results for $\tau \in (0.2, 0.5, 0.8)$ are presented in Table 1 and Table 2 for ease of reading.

All computations are performed with R CRAN software [27]. QRRE, ERRE and PQR were adjusted using respectively quantreg [18], expectreg [32, 31, 29] and rqpd [19] packages.

4.1.2 Results

In order to compare quantile regression estimators to least squares estimators, Koenker [16] decided to focus exclusively on the performance of the median slope estimate. With the expectile regression there is no need to make such restriction, we can compare both methods on any asymmetric value.

In Table 1 we reported the estimated bias and the estimated root-mean-squared error (RMSE) of the different location shift error distributions, and in Table 2 the results for the different location-scale shift error distributions.

The results show that for a fixed T, the bias decreases as n becomes large. The performance indicators (bias and the RMSE) of the different methods are in the same order of magnitude with respect to the different simulation scenarios. We expect poor performance in the location-scale scenario, Table 2. Indeed, in this scenario the effect of the covariate x is:

$$\begin{cases} \beta^Q(\tau) = & \beta + \gamma q(\tau, v) \\ \beta^E(\tau) = & \beta + \gamma \mu(\tau, v) \end{cases}$$

respectively in the quantile regression and expectile regression. For example, in the χ^2_3 distribution the covariate effect for $\tau \in (0.2, 0.5, 0.8)$ is $(0.1, 0.236, 0.464)$ and $(0.191, 0.3, 0.447)$ respectively for the quantile regression and expectile regression. This explains partly the poor performance of the estimators for the χ^2_3 distribution in the location-scale shift scenario in regard to other distributions. By symmetry, the effect of the covariate on the median or the mean is zero in the Normal and t_3 distributions. But adjustment for the true bias has to be made accordingly for $\tau \in (0.1, 0.2, 0.8, 0.9)$. This explain why the performance of the different methods (ERRE, QRRE, PQR) is more noticeable for the median or the mean compared to the other quantiles or expectiles.

In Table 1, the results showed that the quantile regression estimators (QRRE

and PQR) do slightly better than the expectile random effect (ERRE) estimator, but both are competitive. The QRRE estimator performs as well as the PQR estimator, but the penalization is worthwhile. In Table 2, we see that, in general, ERRE is quite competitive with QRRE and PQR except for the χ^2_3 scenario. The performance difference between QRRE and PQR is more remarkable in the location-scale shift scenario (Table 2) where PQR with the Gaussian tuning parameter ($\lambda = \sigma_v/\sigma_u$) performs better.

One of the main advantages of QR and ER methods is their potential to evaluate the influence of factors in several points of the distribution of the dependant variable. In order to emphasise this advantage we display graphically the results of the estimated coefficient and its confidence interval. Figure **??** present the results for the different regression methods (ERRE, QRRE and PQR) in the location shift model, and Figure 3 present the results of the location-scale shift version. The results of the graphics confirm that of the Tables and are easy to read. Overall the different methods are competitive. We can see clearly in the graphics that the ERRE estimate has lower RMSE.

In conclusion, the performance of the quantile regression and the expectile regression are comparable despite their merits and weaknesses. The quantile regression estimates are more robust and the expectile regression estimates are generalization of the mean regression and are easily computable.

4.2 Application

The progress and technological innovation contribute to the creation and disappearance of employment. This upheaval of the labor market spawned a renewed interest in the study of economic returns to education. Koop and Tobias [21] studied this subject with data from the US National Longitudinal Survey of Youth (NLSY). They adjusted a Bayesian hierarchical models to the data in order to evaluate the heterogeneity of returns to schooling. Their results show a presence of heterogeneity and confirm Card assumptions [4], who suggested including a random factor (for the intercept and slope) reflecting the difference between individuals and the heterogeneity of the marginal return to education. QR and ER allow the study of heterogeneity of effects without assuming a prior distribution of the model parameters.

4.2.1 Data

We use Koop and Tobias data for application of the methods, QR and ER, to real data. The data is from the US National Longitudinal Survey of Youth

(NLSY), which began in 1979. The cohort contained initially 12,686 respondents aged between 14 and 22 years. The survey was still ongoing and is in its 25th cycle in 2012. NLSY is an important collection of data that contains information on several topics including education, employment, income, salary and health, among others. In the data cleansing process, the authors excluded some observations because education or wages were unusable. In the end, the database is comprised of Caucasian men aged 16 years (at the beginning of the survey), who report having worked at least 30 weeks or 800 hours per year and earning an hourly wage between 1 and 100$. Data is freely available on the journal website (Journal of Applied Econometrics).

The log of the hourly wage is used as the dependent variable and the variable of interest is the education of the respondents. The other explanatory variables are time invariant characteristics (score on cognitive ability, highest grade completed by the respondent's mother and father, number of siblings, lived or not in a broken home as of age 14) and time varying components (potential labor market experience and its square, and a time trend variable). Their model includes [21] a continuous measure of the local unemployment rate in the given year, which we didn't include in our model due to restricted access. Apart from this variable, our Mincer equation are identical. Summary sample statistics for the selected variable are reported in [21]. Our Mincer equation for both methods QR and ER is:

$$
\begin{aligned}
&q(\tau, \log(\text{Wage})) \text{ or } \mu(\tau, \log(\text{Wage})) = \\
&\quad \beta_{0\tau} + \beta_{1\tau}\text{Education} + \beta_{2\tau}\text{Experience} + \beta_{3\tau}(\text{Experience})^2 + \\
&\quad \beta_{4\tau}(\text{Time trend}) + \beta_{5\tau}\text{Ability} + \beta_{6\tau}(\text{Mother Education}) + \\
&\quad \beta_{7\tau}(\text{Father Education}) + \beta_{8\tau}(\text{Broken Home}) + \beta_{9\tau}\text{Sibling} .
\end{aligned}
$$

4.2.2 Results

We estimated conditional quantiles and expectiles for the series of asymmetric points $(0.05, 0.06, 0.07, \ldots, 0.95)$ of length 91 and generated confidence intervals by bootstrap replications $m = 1000$. The results of time varying variables, according to the method (QR and ER), are presented in Figure 4 and those of time invariant in Figure 5. The use of asymmetric weight indicates the presence of heterogeneity in the economic returns to education. This effect is significantly heterogeneous and increases non-linearly with respect to the degree of asymmetry. The results did not show a significant effect of invariant time variables, except for the ability and the size of the family that appear to have a significant effect on salary, Figure 5.

Our findings are similar to those found in the literature [4] and particularly in [21]. However, we see little differences here and there. For example, our results show that the heterogeneity to return to schooling ranges between 0.0708 and 0.1111 for ERRE and between 0.0612 and 0.1148 for QRRE. While the models in [21] offers a variation of heterogeneity between -0.04 and 0.27.

5 Conclusion

We present the QR and ER estimators for the random effects model. We demonstrate their asymptotic properties and propose an estimator of the variances-covariances matrix. We evaluate their performance in practice through simulations and a real data set.

The simulation results show that the QR estimate does slightly better than the ER estimate but both are competitive. The real data analysis show that the two methods are appropriate to study the heterogeneity of the effects of a factor on the dependent variable. The analysis of the performance of the ER method relative to the penalized quantile regression suggests that the extension of the ER to the penalized fixed effect model is a promising area of research.

Expectiles are a reliable alternative to the quantiles. They can not be interpreted easily but they have an explicit form and are analytically estimable. Furthermore, estimates of their variances-covariances matrix can be evaluated without estimating the error density function. This is not the case for quantiles. Since both methods characterise the c.d.f of the dependent variable, it is not imperative to find each series of quantiles corresponding to a series of expectiles when applying the ER method. And as expectiles are to the mean what quantiles are to the median, they should be used and interpreted as quantilized mean.

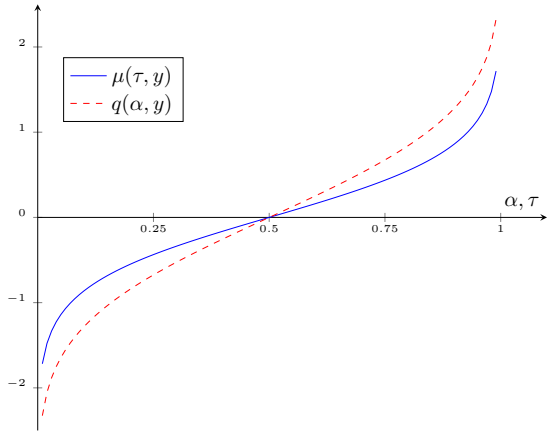

Figure 1: Quantile and expectile of the standard normal distribution $\mathcal{N}(0, 1)$.

Table 1: Bias and root mean squared error (RMSE) for the ERRE, QRRE and PQR models with different location shift error distributions.

		$T=5$			$T=15$			$T=25$		
	τ	50	100	500	50	100	500	50	100	500
$\varepsilon \sim \mathcal{N}(0,1)$										
ERRE	0.2	0.0372	0.0165	0.0389	0.0398	0.0239	-0.0201	-0.0586	-0.0398	-0.0263
		(0.0699)	(0.056)	(0.0289)	(0.0823)	(0.0536)	(0.0275)	(0.0552)	(0.0449)	(0.022)
	0.5	0.0194	0.0328	0.0372	0.0755	0.0461	-0.0084	-0.0703	-0.0327	-0.019
		(0.0705)	(0.0563)	(0.0284)	(0.0751)	(0.0525)	(0.0253)	(0.0545)	(0.0421)	(0.0212)
	0.8	-0.002	0.0325	0.0245	0.1137	0.0748	2e-04	-0.0809	-0.0203	-0.0154
		(0.0827)	(0.0624)	(0.0305)	(0.0764)	(0.0528)	(0.0271)	(0.0536)	(0.0414)	(0.0228)
QRRE	0.2	0.0321	0.0088	0.0253	-0.0088	0.004	-0.0159	-0.038	-0.0388	-0.0297
		(0.083)	(0.0798)	(0.0275)	(0.0832)	(0.0602)	(0.0336)	(0.0632)	(0.0494)	(0.025)
	0.5	0.0239	0.0749	0.0511	0.0634	0.0322	-0.0044	-0.0581	-0.0462	-0.0101
		(0.0858)	(0.0643)	(0.0321)	(0.084)	(0.0552)	(0.0253)	(0.0644)	(0.0388)	(0.023)
	0.8	0.0373	0.0385	9e-04	0.1656	0.0833	0.0111	-0.0976	-0.0208	-0.0092
		(0.0946)	(0.0598)	(0.0407)	(0.0977)	(0.062)	(0.0276)	(0.0635)	(0.0435)	(0.0272)
PQR	0.2	0.028	-5e-04	0.0224	-0.0162	0.0049	-0.01	-0.022	-0.0361	-0.0305
		(0.0855)	(0.0666)	(0.0283)	(0.0855)	(0.0615)	(0.0353)	(0.0507)	(0.0493)	(0.0252)
	0.5	0.0081	0.061	0.0475	0.055	0.0357	9e-04	-0.0433	-0.0454	-0.0112
		(0.086)	(0.071)	(0.0342)	(0.0784)	(0.0614)	(0.0267)	(0.0637)	(0.0423)	(0.022)
	0.8	0.0122	0.0355	-5e-04	0.1458	0.0894	0.0143	-0.0897	-0.0245	-0.0098
		(0.102)	(0.0691)	(0.0379)	(0.0965)	(0.0616)	(0.0298)	(0.0702)	(0.0458)	(0.0272)
$\varepsilon \sim t_3$										
ERRE	0.2	0.2455	-0.0087	0.0651	0.1	0.024	0.0289	0.0685	0.0378	0.0325
		(0.153)	(0.1734)	(0.065)	(0.1768)	(0.0538)	(0.056)	(0.0723)	(0.0661)	(0.0359)
	0.5	0.1021	-0.1228	0.0382	0.1964	0.0477	0.0219	0.0796	0.039	0.0202
		(0.1333)	(0.1301)	(0.0516)	(0.2144)	(0.0512)	(0.04)	(0.0823)	(0.0693)	(0.0335)
	0.8	-0.0492	-0.3222	0.0019	0.3878	0.0778	0.0315	0.1315	0.0597	0.0214
		(0.138)	(0.2474)	(0.0716)	(0.3343)	(0.0504)	(0.0421)	(0.12)	(0.0938)	(0.0422)
QRRE	0.2	0.3081	-0.0976	0.0747	-0.0364	0.0086	0.04	0.0562	0.0374	0.0389
		(0.1388)	(0.128)	(0.0489)	(0.1696)	(0.0624)	(0.0431)	(0.0839)	(0.0601)	(0.0332)
	0.5	0.112	0.0251	0.0498	0.0599	0.0334	-0.0083	0.0507	0.0049	0.0011
		(0.1199)	(0.0909)	(0.0366)	(0.1202)	(0.0582)	(0.0293)	(0.0878)	(0.0722)	(0.0332)
	0.8	-0.174	-0.2089	0.0492	0.2207	0.0814	0.021	0.0381	0.0086	-0.0063
		(0.2359)	(0.1327)	(0.0615)	(0.178)	(0.0651)	(0.0337)	(0.0918)	(0.106)	(0.0436)
PQR	0.2	0.2924	-0.1163	0.0729	-0.024	0.0095	0.0443	0.0673	0.0416	0.0428
		(0.1498)	(0.1268)	(0.0518)	(0.1606)	(0.0634)	(0.0395)	(0.079)	(0.0637)	(0.0346)
	0.5	0.1199	0.0191	0.0498	0.0641	0.0393	-0.0042	0.0514	0.0162	0.0044
		(0.1249)	(0.0909)	(0.0359)	(0.1221)	(0.0623)	(0.0311)	(0.087)	(0.0668)	(0.0307)
	0.8	-0.1614	-0.2213	0.042	0.2226	0.0887	0.0294	0.0471	0.0106	-0.002
		(0.2255)	(0.1378)	(0.0631)	(0.1777)	(0.0642)	(0.0362)	(0.0984)	(0.1036)	(0.0392)

Table 1: Bias and root mean squared error (RMSE) for the ERRE, QRRE and PQR models with different location shift error distributions.

		$T = 5$			$T = 15$			$T = 25$		
	τ	50	100	500	50	100	500	50	100	500
$\varepsilon \sim \chi^2_3$										
ERRE	0.2	0.2055	0.0803	0.0635	0.3672	0.1662	0.0082	5e-04	0.0346	0.0399
		(0.178)	(0.1405)	(0.055)	(0.1182)	(0.0937)	(0.0456)	(0.1351)	(0.1012)	(0.0387)
	0.5	0.2134	0.0458	0.0969	0.438	0.1823	-0.0021	0.0747	0.0867	0.0815
		(0.22)	(0.1939)	(0.0725)	(0.1307)	(0.1121)	(0.0584)	(0.153)	(0.1152)	(0.0487)
	0.8	0.2997	0.136	0.12	0.5093	0.2433	-0.031	0.1878	0.1745	0.1342
		(0.2872)	(0.261)	(0.1007)	(0.1592)	(0.1605)	(0.0792)	(0.2112)	(0.1529)	(0.0634)
QRRE	0.2	0.2054	0.0887	0.0196	0.3119	0.1516	0.0106	-0.1255	-0.0301	-0.0041
		(0.1937)	(0.1112)	(0.0596)	(0.1449)	(0.1189)	(0.045)	(0.1721)	(0.1051)	(0.0429)
	0.5	0.0732	-0.1211	0.1107	0.4613	0.1351	0.042	-0.0099	0.0308	0.0458
		(0.2308)	(0.2067)	(0.0666)	(0.134)	(0.1078)	(0.0627)	(0.1862)	(0.1254)	(0.0595)
	0.8	0.3619	-0.1012	0.1583	0.4821	0.2339	-0.0878	0.135	0.155	0.2123
		(0.4064)	(0.3431)	(0.1502)	(0.2019)	(0.2389)	(0.0959)	(0.2539)	(0.1445)	(0.0998)
PQR	0.2	0.2681	0.0916	0.022	0.2493	0.1414	0.0223	-0.1407	-0.0223	0.0024
		(0.1746)	(0.1087)	(0.0566)	(0.1384)	(0.1202)	(0.0443)	(0.1681)	(0.1007)	(0.0449)
	0.5	0.1107	-0.0976	0.1305	0.3664	0.1245	0.0529	-0.001	0.0487	0.0515
		(0.2604)	(0.1966)	(0.0709)	(0.1455)	(0.1)	(0.0608)	(0.193)	(0.1232)	(0.0601)
	0.8	0.3827	-0.0633	0.1524	0.5016	0.205	-0.0777	0.1704	0.1523	0.221
		(0.4016)	(0.3464)	(0.1566)	(0.1796)	(0.2198)	(0.0895)	(0.2722)	(0.1338)	(0.0892)

Table 2: Bias and root mean squared error (RMSE) for the ERRE, QRRE and PQR models with different location-scale shift error distributions.

		$T=5$			$T=15$			$T=25$		
	τ	50	100	500	50	100	500	50	100	500
$\varepsilon \sim \mathcal{N}(0,1)$										
ERRE	0.2	-0.0338	0.0162	-0.0526	-0.0224	-0.0656	-0.0528	-0.0758	-0.0759	-0.0425
		(0.0509)	(0.0372)	(0.0161)	(0.0414)	(0.0262)	(0.0127)	(0.029)	(0.0202)	(0.0099)
	0.5	0.0017	0.0658	-0.0137	0.0317	-0.0092	-0.007	-0.0264	-0.0231	0.0034
		(0.0402)	(0.031)	(0.0152)	(0.0346)	(0.0268)	(0.0115)	(0.0299)	(0.0176)	(0.0095)
	0.8	0.0525	0.1055	0.0295	0.0859	0.0445	0.0368	0.0189	0.0313	0.0505
		(0.0424)	(0.0296)	(0.0158)	(0.0339)	(0.0277)	(0.0117)	(0.0355)	(0.0187)	(0.0096)
QRRE	0.2	-0.0246	3e-04	-0.0709	-0.0499	-0.1015	-0.0821	-0.1202	-0.109	-0.066
		(0.0707)	(0.051)	(0.0173)	(0.0472)	(0.0336)	(0.0156)	(0.0371)	(0.0283)	(0.0105)
	0.5	-0.0276	0.0681	-0.0243	0.0246	-0.0077	0	-0.0384	-0.0327	-4e-04
		(0.0343)	(0.0398)	(0.0172)	(0.0289)	(0.033)	(0.012)	(0.0306)	(0.0163)	(0.0109)
	0.8	0.0666	0.1381	0.0395	0.0923	0.078	0.0653	0.0295	0.047	0.0834
		(0.0655)	(0.0355)	(0.0177)	(0.0421)	(0.0312)	(0.0144)	(0.0483)	(0.0223)	(0.0119)
PQR	0.2	-0.0235	-0.0115	-0.069	-0.0598	-0.1002	-0.0801	-0.1078	-0.1049	-0.0664
		(0.0686)	(0.0513)	(0.0182)	(0.0438)	(0.0306)	(0.0149)	(0.043)	(0.0284)	(0.0101)
	0.5	-0.0343	0.06	-0.0234	0.0128	-0.0087	-9e-04	-0.0277	-0.031	-0.0013
		(0.0341)	(0.0357)	(0.0186)	(0.033)	(0.031)	(0.0132)	(0.0352)	(0.017)	(0.0107)
	0.8	0.052	0.1286	0.0395	0.0836	0.0773	0.0646	0.0412	0.0462	0.0824
		(0.0706)	(0.0378)	(0.0184)	(0.0423)	(0.0302)	(0.015)	(0.0451)	(0.0233)	(0.0115)
$\varepsilon \sim t_3$										
ERRE	0.2	-0.071	0.0603	-0.138	-0.0107	-0.0676	-0.0644	-0.0937	-0.0856	-0.0796
		(0.0759)	(0.0842)	(0.0437)	(0.0798)	(0.0265)	(0.018)	(0.0373)	(0.0417)	(0.0157)
	0.5	0.0318	0.0058	-0.0656	0.0916	-0.011	0.0051	-0.0236	-0.0098	-0.0185
		(0.0621)	(0.0995)	(0.0269)	(0.1097)	(0.027)	(0.0151)	(0.0335)	(0.0368)	(0.0139)
	0.8	0.1321	-0.0974	-0.0167	0.2273	0.0434	0.0795	0.0467	0.0461	0.0437
		(0.0831)	(0.2086)	(0.0287)	(0.1532)	(0.0274)	(0.0182)	(0.0434)	(0.0393)	(0.0165)
QRRE	0.2	-0.0665	0.0503	-0.1423	-0.0779	-0.1051	-0.0822	-0.0976	-0.0832	-0.1027
		(0.1062)	(0.0454)	(0.0205)	(0.0623)	(0.0319)	(0.0212)	(0.0405)	(0.032)	(0.0164)
	0.5	0.021	0.0365	-0.0338	0.0316	-0.0099	0.0119	-0.0118	0.0145	-0.0229
		(0.0713)	(0.0512)	(0.018)	(0.069)	(0.0303)	(0.0135)	(0.0394)	(0.0293)	(0.0121)
	0.8	0.1846	0.1096	0.0353	0.236	0.0792	0.0895	0.0743	0.0645	0.0469
		(0.1014)	(0.0839)	(0.0237)	(0.159)	(0.0297)	(0.0216)	(0.0413)	(0.0413)	(0.0146)
PQR	0.2	-0.0645	0.046	-0.1411	-0.079	-0.1051	-0.0825	-0.1045	-0.0796	-0.1052
		(0.1079)	(0.046)	(0.0215)	(0.0599)	(0.0314)	(0.0206)	(0.0411)	(0.0316)	(0.018)
	0.5	0.0311	0.0361	-0.0338	0.0299	-0.0134	0.0118	-0.0144	0.0197	-0.0235
		(0.0737)	(0.0472)	(0.0189)	(0.0719)	(0.0304)	(0.0133)	(0.0412)	(0.0288)	(0.0119)
	0.8	0.1925	0.1116	0.0337	0.2295	0.0738	0.0893	0.0772	0.0705	0.0466
		(0.0974)	(0.076)	(0.0245)	(0.1539)	(0.0312)	(0.0208)	(0.0419)	(0.0384)	(0.0134)

Table 2: Bias and root mean squared error (RMSE) for the ERRE, QRRE and PQR models with different location-scale shift error distributions.

		$T = 5$			$T = 15$			$T = 25$		
	τ	50	100	500	50	100	500	50	100	500
$\varepsilon \sim \chi^2_3$										
ERRE	0.2	0.4884	0.131	0.1862	0.2131	0.2196	0.2226	0.2259	0.2427	0.2332
		(0.1369)	(0.0665)	(0.0278)	(0.0548)	(0.0386)	(0.021)	(0.0567)	(0.0356)	(0.0182)
	0.5	0.6781	0.2161	0.2834	0.3328	0.282	0.3123	0.3209	0.3375	0.3249
		(0.1428)	(0.0946)	(0.0379)	(0.0731)	(0.0517)	(0.0272)	(0.0723)	(0.0431)	(0.0218)
	0.8	0.8936	0.3229	0.4119	0.4926	0.3816	0.4398	0.4439	0.4779	0.4461
		(0.1737)	(0.1476)	(0.0544)	(0.103)	(0.0682)	(0.0361)	(0.0953)	(0.0601)	(0.0286)
QRRE	0.2	0.2112	0.0418	0.1045	0.1336	0.1858	0.1564	0.1354	0.1849	0.1667
		(0.1741)	(0.0583)	(0.0254)	(0.0558)	(0.0349)	(0.0212)	(0.0614)	(0.0387)	(0.0206)
	0.5	0.5525	0.1528	0.2496	0.3119	0.2357	0.2527	0.2648	0.2561	0.2764
		(0.1717)	(0.1011)	(0.0485)	(0.0864)	(0.0462)	(0.0284)	(0.0956)	(0.0479)	(0.0256)
	0.8	0.8481	0.3288	0.4171	0.5088	0.3688	0.4653	0.4136	0.4378	0.4597
		(0.177)	(0.1531)	(0.074)	(0.1077)	(0.0763)	(0.0486)	(0.099)	(0.0661)	(0.0309)
PQR	0.2	0.2605	0.0371	0.0992	0.084	0.1853	0.1548	0.1335	0.1835	0.1628
		(0.1697)	(0.0542)	(0.0234)	(0.0524)	(0.0387)	(0.0211)	(0.0601)	(0.0381)	(0.0212)
	0.5	0.5687	0.1644	0.2346	0.2653	0.2327	0.2531	0.2694	0.2564	0.2737
		(0.1723)	(0.1014)	(0.0485)	(0.0953)	(0.0503)	(0.0304)	(0.1039)	(0.0493)	(0.0272)
	0.8	0.8645	0.3399	0.4059	0.4633	0.3717	0.468	0.4287	0.4409	0.4573
		(0.1747)	(0.1531)	(0.0773)	(0.1053)	(0.0744)	(0.0472)	(0.1174)	(0.0642)	(0.0302)

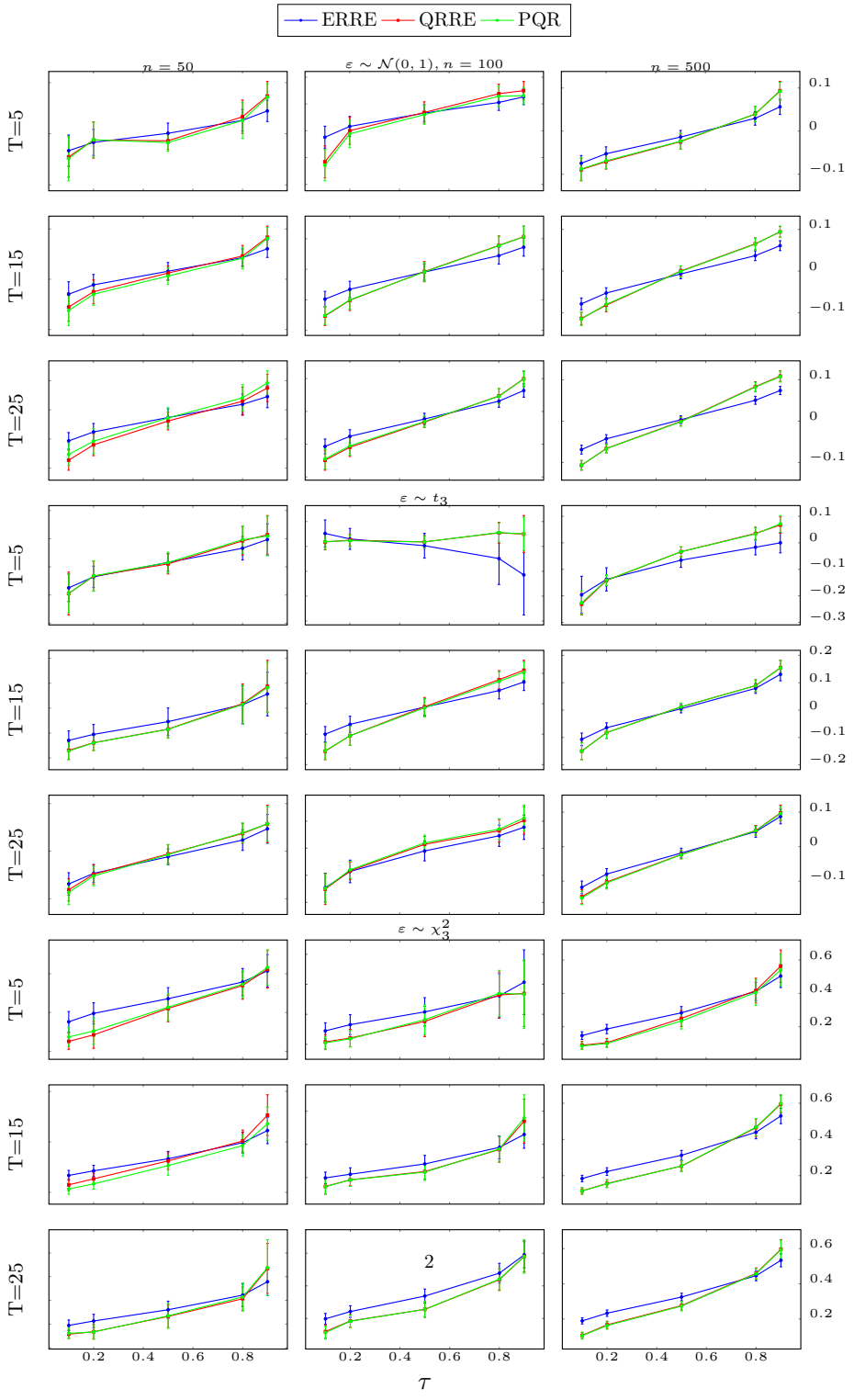

Figure 2: Estimated coefficient and confidence interval for the different regression methods in the location-scale shift model.

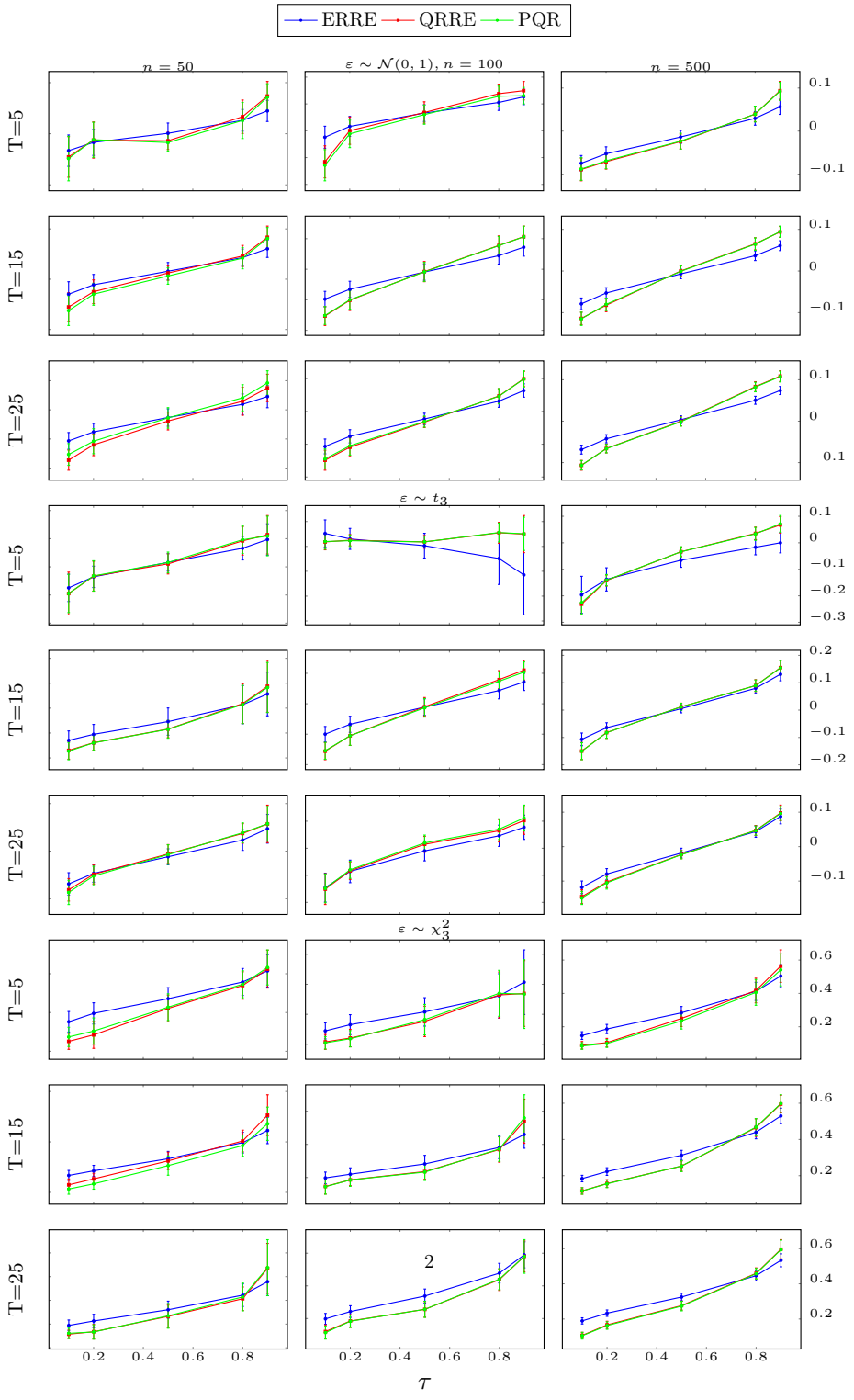

Figure 3: Estimated coefficient and confidence interval for the different regression methods in the location-scale shift model.

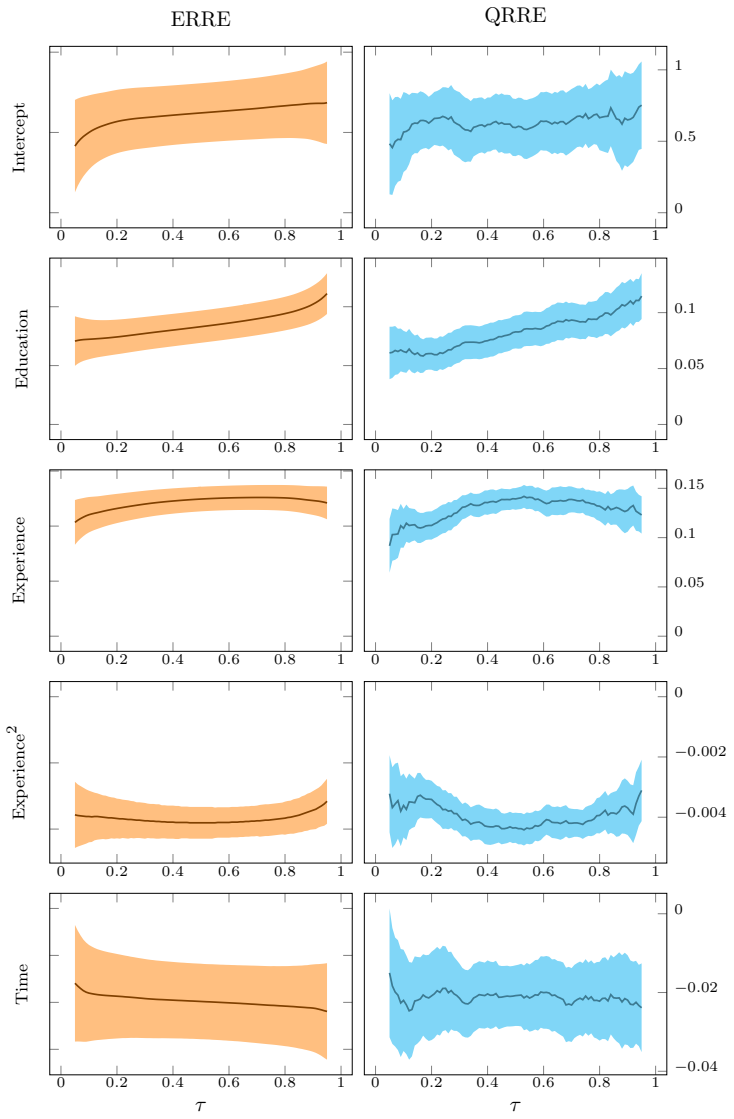

Figure 4: Quantile and expectile estimated coefficients of time-varying characteristics with nonparametric confident interval.

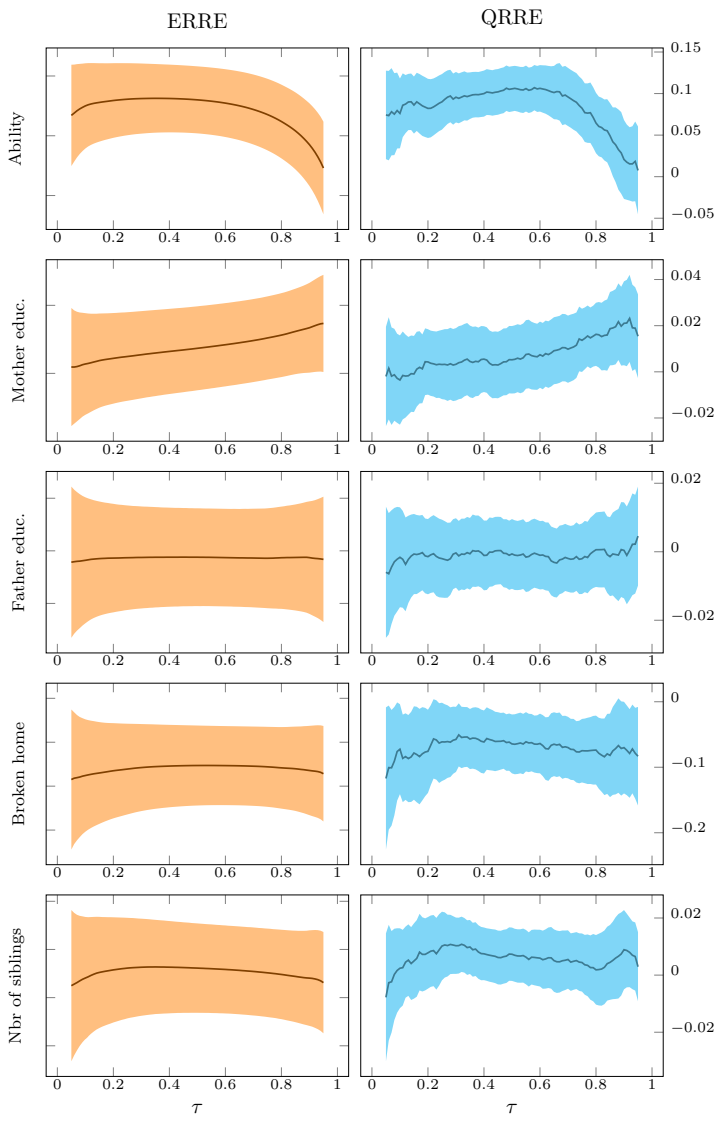

Figure 5: Quantile and expectile estimated coefficients of time-invariant characteristics with nonparametric confident interval.

6 Proofs

Notice that in this section the expectations are conditional to $\boldsymbol{X}$. As the proofs on the quantiles and the expectiles are separated we omit the exponents (Q and E) on $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$.

Proof of Theorem 3.1.

The proof follows closely that of [**?**, **?**]. Consider the new objective function

$$Z_N(\delta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\Big[r_\alpha\big\{\varepsilon_{it}(\alpha) - \boldsymbol{x}_{it}{}^{\mathsf{T}}\delta/\sqrt{N}\big\} - r_\alpha\{\varepsilon_{it}(\alpha)\}\Big].$$

This function $\{Z_N(\delta)\}$ and the initial objective function $Q_N\{\boldsymbol{\beta}(\alpha), \alpha\}$ have the same extremum. $Z_N(\delta)$ is a sum of convex functions and admits as minimum $\widehat{\delta}_N = \sqrt{N}\{\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)\}$. The idea of the proof is to give an approximation of $Z_N(\delta)$ by a quadratic function, and to show that $\widehat{\delta}_N$ has the same asymptotic properties as the extreme value of this quadratic function. The following identity ([17], p.121) provides this approximation.

$$r_\alpha(\varepsilon - \upsilon) - r_\alpha(\varepsilon) = -\upsilon\psi_\alpha(\varepsilon) + \int_0^\upsilon \Big\{\mathbb{1}(\varepsilon < s) - \mathbb{1}(\varepsilon \leq 0)\Big\}ds, \tag{12}$$

with $\psi_\alpha(\varepsilon) = \alpha - \mathbb{1}(\varepsilon < 0)$. From this identity (12), the new risk function is divided in two functions, $Z_N(\delta) = Z_{1N}(\delta) + Z_{2N}(\delta)$, where:

$$\begin{aligned} Z_{1N}(\delta) &= -\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta\psi_\alpha\{\varepsilon_{it}(\alpha)\} \\ &= -\frac{1}{\sqrt{N}}\sum_{i=1}^{n}\delta^{\mathsf{T}}\boldsymbol{X}_i{}^{\mathsf{T}}\psi_\alpha\{\boldsymbol{\varepsilon_i}(\alpha)\} \end{aligned}$$

and

$$\begin{aligned} Z_{2N}(\delta) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\upsilon_{Nit}}\Big[\mathbb{1}\{\varepsilon_{it}(\alpha) < s\} - \mathbb{1}\{\varepsilon_{it}(\alpha) \leq 0\}\Big]ds \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T}Z_{2Nit}(\delta) \end{aligned}$$

with $\quad \upsilon_{Nit} = \boldsymbol{x}_{it}{}^{\mathsf{T}}\delta/\sqrt{N}, \quad$ and $\quad \psi_\alpha\{\boldsymbol{\varepsilon_i}(\alpha)\} = \Big[\psi_\alpha\{\varepsilon_{i1}(\alpha)\}, \ldots, \psi_\alpha\{\varepsilon_{iT}(\alpha)\}\Big]^{\mathsf{T}}.$

The first and second moment of the component of the random vector $\psi_\alpha\{\boldsymbol{\varepsilon_i}(\alpha)\}$ are:

$$\mathbb{E}\left[\psi_\alpha\{\varepsilon_{it}(\alpha)\}\right] = 0$$

$$\mathrm{Var}\left[\psi_\alpha\{\varepsilon_{it}(\alpha)\}\right] = \alpha(1-\alpha)$$

$$\mathrm{Cov}\left[\psi_\alpha\{\varepsilon_{it}(\alpha)\}, \psi_\alpha\{\varepsilon_{is}(\alpha)\}\right] = \sigma_{its}(\alpha) = \mathbb{E}\left[\mathbb{1}\big\{\varepsilon_{it}(\alpha) < 0, \varepsilon_{is}(\alpha) < 0\big\}\right] - \alpha^2.$$

Then by the Lindeberg-Feller central limit theorem, with condition **Q2** it can be deduced that $Z_{1N}(\delta) \xrightarrow{d} -\delta^{\mathsf{T}}W$ where $W \sim \mathcal{N}(0, \mathbf{D}_0)$.

The second term of $Z_N(\delta)$ can be rewritten:

$$Z_{2N}(\delta) = \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left[Z_{2Nit}(\delta)\right] + \sum_{i=1}^{n}\sum_{t=1}^{T}\Big[Z_{2Nit}(\delta) - \mathbb{E}\left\{Z_{2Nit}(\delta)\right\}\Big].$$

We show that:

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left\{Z_{2Nit}(\delta)\right\} &= \sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\upsilon_{Nit}}\mathbb{E}\Big[\mathbb{1}\{\varepsilon_{it}(\alpha)<s\}-\mathbb{1}\{\varepsilon_{it}(\alpha)\leq 0\}\Big]ds \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\upsilon_{Nit}}\Big[F_{it}\{q(\alpha,y_{it})+s\}-F_{it}\{q(\alpha,y_{it})\}\Big]ds \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta}\frac{1}{\sqrt{N}}\Big[F_{it}\{q(\alpha,y_{it})+s/\sqrt{N}\}-F_{it}\{q(\alpha,y_{it})\}\Big]ds \\
&= N^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta}\sqrt{N}\Big[F_{it}\{q(\alpha,y_{it})+s/\sqrt{N}\}-F_{it}\{q(\alpha,y_{it})\}\Big]ds \\
&= N^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\int_0^{\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta}f_{it}\{q(\alpha,y_{it})\}sds+o(1) \\
&= (2N)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}f_{it}\{q(\alpha,y_{it})\}\delta^{\mathsf{T}}\boldsymbol{x}_{it}\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta+o(1) \\
&\to \frac{1}{2}\delta^{\mathsf{T}}\mathbf{D}_1\delta.
\end{aligned}
\tag{13}
$$

The third line of the above equation (13) is obtained through substitution of s by $s/\sqrt{N}$ and by multiplying by $\sqrt{N}$ the third line from the bottom follows by condition **Q1**. Now noticing that, with condition **Q2**.3:

$$
\mathrm{Var}\{Z_{2N}(\delta)\}\leq\frac{1}{N}\max\left\|\boldsymbol{x}_{it}{}^{\mathsf{T}}\delta\right\|\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\{Z_{2Nit}(\delta)\}\to 0
$$

and applying the Chebychev's Inequality, the remaining term converge to 0 in probability:

$$
Z_{2N}(\delta)-\mathbb{E}\left\{Z_{2N}(\delta)\right\}\xrightarrow{\mathbb{P}}0. \tag{14}
$$

From (13) and (14), we obtain: $Z_{2N}(\delta)\xrightarrow{\mathbb{P}}\frac{1}{2}\delta^{\mathsf{T}}\mathbf{D}_1\delta$. Now from the properties of the first term $Z_{1N}(\delta)$ and the second term $Z_{2N}(\delta)$ we can deduce that:

$$
Z_N(\delta)\xrightarrow{d}Z_0(\delta)=-\delta^{\mathsf{T}}W+\frac{1}{2}\delta^{\mathsf{T}}\mathbf{D}_1\delta.
$$

$Z_0(\delta)$ has a unique minimum $\mathbf{D}_1^{-1}W$. Now the two convexity lemmas of [**?**] state that if a random convex functions $Z_N(\delta)$ converge in distribution to some function $Z_0(\delta)$ which has a unique minimum $\mathbf{D}_1^{-1}W$ then $\widehat{\delta}_N\xrightarrow{d}\mathbf{D}_1^{-1}W$. Hence we conclude that:

$$\sqrt{N}\{\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)\} \xrightarrow{d} \mathcal{N}(0, \mathbf{D}_1^{-1}\mathbf{D}_0\mathbf{D}_1^{-1}).$$

□

Proof of Theorem 3.2.

Without lost of generality, let F_N and F be respectively the joint empirical and population distribution function of (y_{it}, y_{is}).

By Lemma A1 of [**?**], we have: $F_N\{\boldsymbol{x}_{it}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha)\} \xrightarrow{\mathbb{P}} F\{\boldsymbol{x}_{it}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha)\}$ uniformly.

With $\widehat{\boldsymbol{\beta}}(\alpha) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}(\alpha)$, we apply Lemma 4 of [**?**], then:

$$F\{\boldsymbol{x}_{it}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha)\} \xrightarrow{\mathbb{P}} F\{\boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\boldsymbol{\beta}(\alpha)\}.$$

Thus:

$$F_N\{\boldsymbol{x}_{it}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\widehat{\boldsymbol{\beta}}(\alpha)\} \xrightarrow{\mathbb{P}} F\{\boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\alpha), \boldsymbol{x}_{is}{}^\mathsf{T}\boldsymbol{\beta}(\alpha)\}.$$

Hence $\widehat{\sigma}_{its}(\alpha) \xrightarrow{\mathbb{P}} \sigma_{its}(\alpha)$, and by the Slutsky theorem $\widehat{\mathbf{D}}_1\widehat{\mathbf{D}}_0\widehat{\mathbf{D}}_1 \xrightarrow{\mathbb{P}} \mathbf{D}_1^{-1}\mathbf{D}_0\mathbf{D}_1^{-1}$.

□

Proof of Theorem 3.3.

Let $R\{\boldsymbol{\beta}(\tau),\tau\} = \mathbb{E}[h_i\{\boldsymbol{\beta}(\tau)\}]$, with $h_i\{\boldsymbol{\beta}(\tau)\} = \sum_{t=1}^{T} r_\tau^E\{y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\}$ and,

$$\begin{aligned} R_N\{\boldsymbol{\beta}(\tau),\tau\} &= N^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} r_\tau^E\{y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\} \\ &= N^{-1}\sum_{i=1}^{n} h_i\{\boldsymbol{\beta}(\tau)\} \end{aligned}$$

The proof follows closely that of [25]. First we show the consistency of $\widehat{\boldsymbol{\beta}}(\tau)$ with Lemma A of [25]. The application of the second order sufficient optimality condition allows us to have the uniqueness of $\boldsymbol{\beta}(\tau)$. Lemma A1 of [**?**] allows us to obtain the uniform convergence in probability of $R_N\{\boldsymbol{\beta}(\tau),\tau\}$ to $R\{\boldsymbol{\beta}(\tau),\tau\}$. Finally, with the convexity of the risk function $R_N\{\boldsymbol{\beta}(\tau),\tau\}$ we have the consistency result.

$h_i\{\boldsymbol{\beta}(\tau)\}$ is sum of convex and differentiable functions in $\boldsymbol{\beta}(\tau)$. Then $h_i\{\boldsymbol{\beta}(\tau)\}$ is derivable and its derivative, $g_i\{\boldsymbol{\beta}(\tau)\}$, is bounded by an integrable function, by assumption **E2**.

Indeed:

$$\begin{aligned} g_i\{\boldsymbol{\beta}(\tau)\} &= \frac{\partial h_i\{\boldsymbol{\beta}(\tau)\}}{\partial \boldsymbol{\beta}} \\ &= \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\beta}} r_\tau^E\{y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\} \\ &= -2\sum_{t=1}^{T} \boldsymbol{x}_{it}\psi_\tau\{y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\} \\ &= -2\boldsymbol{X}_i{}^\mathsf{T} W_i(\tau)\boldsymbol{\varepsilon_i}(\tau). \end{aligned}$$

With $\psi_\tau\{y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\} = |\tau - \mathbb{1}\{y_{it} < \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\}|\Big(y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau)\Big)$ and $W_i(\tau) = \mathrm{Diag}\{w_{i1}(\tau),\ldots,w_{iT}(\tau)\}$. Then we have:

$$\begin{aligned} \|g_i\{\boldsymbol{\beta}(\tau)\}\| &= 2\Big\|\sum_{t=1}^{T} \boldsymbol{x}_{it}\psi_\tau(y_{it} - \boldsymbol{x}_{it}{}^\mathsf{T}\boldsymbol{\beta}(\tau))\Big\| \\ &\leq 2T\,\|z\|^2\,(d + d'\,\|\boldsymbol{\beta}(\tau)\|) \end{aligned}$$

By the dominated convergence theorem of Lebesgue, we can interchange derivative and integral sign. We have:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\beta}} R\{\boldsymbol{\beta}(\tau), \tau\} &= \mathbb{E}[g_i\{\boldsymbol{\beta}(\tau)\}] \\ &= -2\sum_{t=1}^{T} \boldsymbol{x}_{it}\left[\tau \int_{\boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)}^{+\infty} \{y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)\} f(y_{it}) dy_{it}\right. \\ &\quad \left. + (1-\tau)\int_{-\infty}^{\boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)} \{y_{it} - \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)\} f(y_{it}) dy_{it}\right]\end{aligned}$$

By Leibniz integral rule [?], $\int_{-\infty}^{\alpha}(y-\alpha)f(y|x)dy$ is continuously differentiable in α, with derivative $-\int_{-\infty}^{\alpha} f(y|x)dy$ bounded by 1. By the dominated convergence theorem of Lebesgue, $\partial R\{\boldsymbol{\beta}(\tau),\tau\}/\partial\boldsymbol{\beta}$ is continuously differentiable, with derivative:

$$\begin{aligned}\frac{\partial^2 R\{\boldsymbol{\beta}(\tau),\tau\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}} &= 2\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}{}^{\mathsf{T}}\left[\tau\int_{\boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)}^{+\infty} f(y_{it})dy_{it} + (1-\tau)\int_{-\infty}^{\boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)} f(y_{it})dy_{it}\right] \\ &= 2\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}{}^{\mathsf{T}}\,\mathbb{E}\left[\left|\tau - \mathbb{1}\{y_{it} < \boldsymbol{x}_{it}{}^{\mathsf{T}}\boldsymbol{\beta}(\tau)\}\right|\right]\end{aligned}$$

By denoting $\delta = \min\{\tau, (1-\tau)\}$, we show that $\partial^2 R\{\boldsymbol{\beta}(\tau),\tau\}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}$ is a positive definite matrix.

The fact that the function is twice continuously differentiable will allow us to write the Taylor expansion with a second order of the function $R\{\boldsymbol{\beta}(\tau),\tau\}$. With the convexity property, we can show the existence and uniqueness of a global minimum $\boldsymbol{\beta}(\tau)$ of the function $R\{\boldsymbol{\beta}(\tau),\tau\}$.

Let $\widetilde{\boldsymbol{\beta}}(\tau)$ be a point in the neighbour of $\boldsymbol{\beta}(\tau)$, then:

$$\begin{aligned}R\{\boldsymbol{\beta}(\tau),\tau\} - R\{\widetilde{\boldsymbol{\beta}}(\tau),\tau\} &= \left[\partial R\{\widetilde{\boldsymbol{\beta}}(\tau),\tau\}/\partial\boldsymbol{\beta}\right]^{\mathsf{T}}[\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)] \\ &\quad + [\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)]^{\mathsf{T}}\left[\partial^2 R\{\dot{\boldsymbol{\beta}}(\tau),\tau\}/\partial\boldsymbol{\beta}\right][\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)] \\ &\geq \left[\frac{\partial}{\partial\boldsymbol{\beta}} R\{\widetilde{\boldsymbol{\beta}}(\tau),\tau\}\right]^{\mathsf{T}}[\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)] + \delta m_x p|\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)|^2,\end{aligned}$$

$\dot{\boldsymbol{\beta}}(\tau)$ is a point of the segment $[\boldsymbol{\beta}(\tau), \widetilde{\boldsymbol{\beta}}(\tau)]$, p is the number of explicative variables and m_x is the smallest eigenvalues of $\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}{}^{\mathsf{T}}$. If we choose a

$\boldsymbol{\beta}(\tau)$ outside the neighbour of $\widetilde{\boldsymbol{\beta}}(\tau)$ and we divide by $|\boldsymbol{\beta}(\tau) - \widetilde{\boldsymbol{\beta}}(\tau)|$ to cancel the first term of the inequality, we see that $R\{\boldsymbol{\beta}(\tau), \tau\} > R\{\widetilde{\boldsymbol{\beta}}(\tau), \tau\}$. It follows by continuity that the function $R\{\boldsymbol{\beta}(\tau), \tau\}$ has a minimum, $\boldsymbol{\beta}(\tau)$, within the neighbour of $\widetilde{\boldsymbol{\beta}}(\tau)$ and that this minimum is a global minimum. Hence $\boldsymbol{\beta}(\tau)$ is the unique solution of the equation $\mathbb{E}[g_i\{\boldsymbol{\beta}(\tau)\}]$ by the convexity of $R\{\boldsymbol{\beta}(\tau), \tau\}$.

With assumptions **E1** and **E2** $R_N\{\boldsymbol{\beta}(\tau), \tau\}$ function verify condition of Lemma A1 of [**?**]. Thus $R_N\{\boldsymbol{\beta}(\tau), \tau\}$ converge uniformly to $R\{\boldsymbol{\beta}(\tau), \tau\}$ in any compact of $\boldsymbol{\beta}(\tau)$.

We have just shown the uniqueness of $\boldsymbol{\beta}(\tau)$ and the uniform convergence of $R_N\{\boldsymbol{\beta}(\tau), \tau\}$. All conditions of lemma A of [25] are satisfied. We can conclude that $\widehat{\boldsymbol{\beta}}(\tau)$ exist with probability approaching 1 and it converges in probability to the parameter $\boldsymbol{\beta}(\tau)$. In others word, $\widehat{\boldsymbol{\beta}}(\tau)$ is consistent.

Now we prove the asymptotic distribution of the estimators.

For every $\lambda \in \mathbb{R}^p$, let:

$$\begin{aligned} Z_N &= \frac{\lambda^\intercal \partial / \partial \boldsymbol{\beta} R_N\{\boldsymbol{\beta}(\tau), \tau\}}{\left(\lambda^\intercal \operatorname{Var}\left[\partial / \partial \boldsymbol{\beta} R_N\{\boldsymbol{\beta}(\tau), \tau\}\right] \lambda\right)^{1/2}} \\ &= \frac{\sum_{i=1}^n \lambda^\intercal g_i\{\boldsymbol{\beta}(\tau)\}}{\left(\lambda^\intercal \sum_{i=1}^n \operatorname{Var}\left[g_i\{\boldsymbol{\beta}(\tau)\}\right] \lambda\right)^{1/2}} \end{aligned}$$

We have:

$$\begin{aligned} \int \left\|\lambda^\intercal g_i\{\boldsymbol{\beta}(\tau)\}\right\|^2 f(y_i) dy_i &= \int \lambda^\intercal \boldsymbol{X}_i{}^\intercal W_i(\tau) \boldsymbol{\varepsilon_i}(\tau) \boldsymbol{\varepsilon_i}(\tau)^\intercal W_i(\tau) \boldsymbol{X}_i \lambda f(y_i) dy_i \\ &= \sum_{k,l}^{p} \sum_{t,s}^{T} \lambda_k \lambda_l \int x_{it}^k x_{is}^l w_{it}(\tau) w_{is}(\tau) \varepsilon_{it}(\tau) \varepsilon_{is}(\tau) f(y_i) dy_i \\ &\leq p^2 T^2 d \left\|\lambda\right\|^2 \int \left\|\boldsymbol{z}\right\|^4 \alpha(z) < \infty. \end{aligned} \tag{15}$$

With p the number of explicative variables, T the number of repeated observations, d a generic constant and $w_{it}(\tau) = |\tau - \mathbb{1}(\varepsilon_{it}(\tau) < 0)|$. The convergence of the last equation hold by assumption **E3** and for every fixed vector λ.

Let for every $\varepsilon > 0$,

$$A_{N\varepsilon} = \left\{ y : \|\lambda^{\mathsf{T}} g_i\{\boldsymbol{\beta}(\tau)\}\| > \varepsilon \sqrt{\lambda^{\mathsf{T}} \sum_{i=1}^{n} \mathrm{Var}\left[g_i\{\boldsymbol{\beta}(\tau)\}\right]\lambda} \right\}.$$

Since by Chebyshev inequality $A_{N\varepsilon}$ converges to the empty set and with equation (15) we have:

$$\lim_{n\to\infty} \frac{1}{\lambda^{\mathsf{T}} \sum_{i=1}^{n} \mathrm{Var}\left[g_i\{\boldsymbol{\beta}(\tau)\}\right]\lambda} \sum_{i=1}^{n} \int_{A_{N\varepsilon}} \left\|\lambda^{\mathsf{T}} g_i\{\boldsymbol{\beta}(\tau)\}\right\|^2 f(y_i) dy_i = 0$$

Thus the conditions of the Lindeberg-Feller central limit theorem are satisfied, $Z_N \sim \mathcal{N}(0,1)$. And by the Cramer-Wold device,

$$\sqrt{N}\left[\partial R_N\{\boldsymbol{\beta}(\tau),\tau\}/\partial\boldsymbol{\beta}\right] \sim \mathcal{N}(0, 4\Sigma),$$

with:

$$\Sigma = \boldsymbol{X}^{\mathsf{T}}\, \mathbb{E}(W(\tau)\boldsymbol{\varepsilon}(\tau)\boldsymbol{\varepsilon}(\tau)^{\mathsf{T}} W(\tau))\boldsymbol{X}.$$

Now consider the Taylor expansion of the function $R_N\{\boldsymbol{\beta}(\tau),\tau\}$ in the neighbour of $\boldsymbol{\beta}(\tau)$. As $\operatorname{plim} \widehat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}(\tau)$ we have:

$$\begin{aligned}\sqrt{N}\left[\partial R_N\{\widehat{\boldsymbol{\beta}}(\tau),\tau\}/\partial\boldsymbol{\beta}\right] &= \sqrt{N}\left[\partial R_N\{\boldsymbol{\beta}(\tau),\tau\}/\partial\boldsymbol{\beta}\right] \\ &\quad + \left[\partial^2 R_N\{\widetilde{\boldsymbol{\beta}}(\tau),\tau\}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}\right]\sqrt{N}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\}\end{aligned}$$

With $\widetilde{\boldsymbol{\beta}}(\tau)$ which is located in the segment joining $[\widehat{\boldsymbol{\beta}}(\tau), \boldsymbol{\beta}(\tau)]$, and then $\operatorname{plim} \widetilde{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}(\tau)$. Again the assumptions **E1** and **E2** used to apply Lemma A1 of [**?**] and:

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\| \frac{\partial^2 R_N\{\boldsymbol{\beta}(\tau)\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}} - \mathbb{E}\left[\frac{\partial^2 h\{\boldsymbol{\beta}(\tau)\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}}\right] \right\| \xrightarrow{\mathbb{P}} 0.$$

For every compact $\mathcal{B}$ with $\boldsymbol{\beta}(\tau)$ contained in its interior. Applying Lemma 4 of [**?**], we can write:

$$\operatorname{plim} \frac{\partial^2 R_N\{\widetilde{\boldsymbol{\beta}}(\tau)\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}} = \operatorname{plim} \frac{\partial^2 R_N\{\widehat{\boldsymbol{\beta}}(\tau)\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}} = \mathbb{E}\left[\frac{\partial^2 h\{\boldsymbol{\beta}(\tau)\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}}\right].$$

Then,

$$\sqrt{N}[\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)] = -H^{-1}\sqrt{N}\frac{\partial R_N\{\boldsymbol{\beta}(\tau), \tau\}}{\partial \boldsymbol{\beta}}.$$

Hence:

$$\sqrt{N}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\} \xrightarrow{d} \mathcal{N}(0, H^{-1}\Sigma H^{-1})$$

□

Proof of Theorem 3.4.

We have:

$$\begin{aligned}\left|\widehat{H} - H\right| \leq & \left| \sum_{i=1}^{n}\sum_{t=1}^{T} \widehat{w}_{it}(\tau)\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}/N - \sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}(\tau)\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}/N \right| \\ & + \left| \sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}(\tau)\boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}/N - H \right| \quad (16)\end{aligned}$$

The first term after the inequality of equation (16) is bounded by the function $\sum_{i=1}^{n}\sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}^{\mathsf{T}}|\widehat{w}_{it}(\tau) - w_{it}(\tau)|/N$.

We have: $|\widehat{w}_{it}(\tau) - w_{it}(\tau)| = 0$ if

$$\begin{aligned} & y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau) < \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) \text{ or } y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) < \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau) \text{ or} \\ & \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) < \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau) < y_{it} \text{ or } \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau) < \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) < y_{it}\end{aligned}$$

and $|\widehat{w}_{it}(\tau) - w_{it}(\tau)| = |2\tau - 1|$ if

$$\boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau) < y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) \text{ or } \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) < y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}(\tau).$$

Thus $|\widehat{w}_{it}(\tau) - w_{it}(\tau)| = |2\tau - 1|\mathbb{1}\Big(|\varepsilon_{it}(\tau)| < |\boldsymbol{x}_{it}^{\mathsf{T}}(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau))|\Big)$ and as $\operatorname{plim}\widehat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}(\tau)$, we have for every $\varepsilon > 0$,

$$\begin{aligned}|\widehat{w}_{it}(\tau) - w_{it}(\tau)| &= |2\tau - 1|\mathbb{1}\Big(|\varepsilon_{it}(\tau)| < p|\boldsymbol{x}_{it}|\varepsilon\Big) \\ &= |2\tau - 1|\mathbb{1}\Big(\boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) - p|\boldsymbol{x}_{it}|\varepsilon < y_{it} < \boldsymbol{x}_{it}^{\mathsf{T}}\boldsymbol{\beta}(\tau) + p|\boldsymbol{x}_{it}|\varepsilon\Big).\end{aligned}$$

Hence:

$$\left| \sum_{i=1}^{n} \sum_{t=1}^{T} \widehat{w}_{it}(\tau) \boldsymbol{x}_{it} \boldsymbol{x}_{it}^{\mathsf{T}} / N - \sum_{i=1}^{n} \sum_{t=1}^{T} w_{it}(\tau) \boldsymbol{x}_{it} \boldsymbol{x}_{it}^{\mathsf{T}} / N \right|$$
$$\leq \sum_{i=1}^{n} \sum_{t=1}^{T} |\boldsymbol{x}_{it}|^2 \times \mathbb{1}\Big(|\varepsilon_{it}(\tau)| < |\boldsymbol{x}_{it}| \varepsilon \Big) \leq |\boldsymbol{x}_{it}|^2 \alpha_{\varepsilon}(\boldsymbol{x}_{it}) + \varepsilon$$

With $\alpha_{\varepsilon}(\boldsymbol{x}_{it}) = \int_{I(\boldsymbol{x}_{it}, \varepsilon)} \alpha(z) dy$, et $I(\boldsymbol{x}_{it}, \varepsilon) = [\boldsymbol{x}_{it}^{\mathsf{T}} \boldsymbol{\beta}(\tau) - \varepsilon, \boldsymbol{x}_{it}^{\mathsf{T}} \boldsymbol{\beta}(\tau) + \varepsilon]$. $\alpha_{\varepsilon}(\boldsymbol{x}_{it})$ is an increasing and bounded function, thus converges to zero by the theorem of monotone convergence. The first term of the inequality (16) converge to zero in proba. For the second term of the inequality (16), it suffices to note that $|w_{it}(\tau) \boldsymbol{x}_{it} \boldsymbol{x}_{it}^{\mathsf{T}}| \leq \|\boldsymbol{z}\|^2$, and use the hypothesis **E2** which allows us to apply Lemma A1 [**?**] to conclude that $\operatorname{plim} \widehat{H} = H$.

To show consistency of $\widehat{\Sigma}$, we have $|\boldsymbol{x}_{it} w_{it}(\tau) \varepsilon_{it}(\tau) w_{is}(\tau) \varepsilon_{is}(\tau) \boldsymbol{x}_{is}^{\mathsf{T}}| \leq \|\boldsymbol{z}\|^4 M$, with M constant. This result and the assumption **E2** allows to apply again Lemma A1 of [**?**] to show that

$$\frac{1}{N} \sum_{i=1}^{n} g_i\{\boldsymbol{\beta}(\tau)\} g_i\{\boldsymbol{\beta}(\tau)\}^{\mathsf{T}} = \frac{1}{N} \sum_{i=1}^{n} \boldsymbol{X}_i^{\mathsf{T}} W_i(\tau) \boldsymbol{\varepsilon}(\tau) \boldsymbol{\varepsilon}(\tau)^{\mathsf{T}} W_i(\tau) \boldsymbol{X}_i$$
$$\xrightarrow{\mathbb{P}} \boldsymbol{X}^{\mathsf{T}} \mathbb{E}[W(\tau) \boldsymbol{\varepsilon}(\tau) \boldsymbol{\varepsilon}(\tau)^{\mathsf{T}} W(\tau)] \boldsymbol{X}.$$

This result and $\widehat{\boldsymbol{\beta}}(\tau) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}(\tau)$ allows to invoke Lemma 4 of [**?**] and to conclude that: $\widehat{\Sigma} \xrightarrow{\mathbb{P}} \Sigma$. □

References

[1] D.J. Aigner, T. Amemiya, and D.J. Poirier. On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, 17(2):377, 1976.

[2] M. Buchinsky. Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources*, 33(1):88–126, 1998.

[3] Ivan A. Canay. A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386, 2011-10-01.

[4] David Card. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. NBER Working Papers 7769, National Bureau of Economic Research, Inc, June 2000.

[5] B. Efron. Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1(1):93–125, 1991.

[6] B. Fitzenberger. The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics*, 82(2):235–287, 1998.

[7] Nicholas A. Furlotte, Eleazar Eskin, and Susana Eyheramendy. Genome-wide association mapping with longitudinal data. *Genetic Epidemiology*, 36(5):463–471, 2012.

[8] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*, volume Analytical methods for social research. Cambridge University Press, New York, 2007.

[9] M. Geraci and M. Bottai. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics*, 8(1):140–154, 2007. cited By (since 1996)45.

[10] Marco Geraci and Matteo Bottai. Linear quantile mixed models. *Statistics and Computing*, 24(3):461–479, 2014.

[11] M. Harding and C. Lamarche. A quantile regression approach for estimating panel data models using instrumental variables. *Economics Letters*, 104(3):133–135, 2009. cited By (since 1996)7.

[12] Cheng Hsiao. Panel data analysis—advantages and challenges. *TEST*, 16(1):1–22, 2007.

[13] Thomas Kneib. Beyond mean regression. *Statistical Modelling*, 13(4):275–303, 2013.

[14] R. Koenker and Y. Bilias. Quantile regression for duration data: A reappraisal of the pennsylvania reemployment bonus experiments. *Empirical Economics*, 26(1):199–220, 2001.

[15] Roger Koenker. When are expectiles percentiles? *Econometric Theory*, 9(3):526–527, 1993.

[16] Roger Koenker. Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89, 2004-10.

[17] Roger Koenker. *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, 2005.

[18] Roger Koenker. *quantreg: Quantile Regression*, 2016. R package version 5.26.

[19] Roger Koenker and Stefan Holst Bache. *rqpd: Regression Quantiles for Panel Data*, 2014. R package version 0.6/r10.

[20] Roger Koenker and Gilbert Bassett, Jr. Regression quantiles. *Econometrica. Journal of the Econometric Society*, 46(1):33–50, 1978.

[21] Gary Koop and Justin L. Tobias. Learning about heterogeneity in returns to schooling. *Journal of Applied Econometrics*, 19(7):827–849, 2004.

[22] Carlos Lamarche. Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, 157(2):396–408, 2010-08.

[23] José A. F. Machado and J. M. C. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.

[24] Yunming Mu Masha Kocherginsky, Xuming He. Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, 14(1):41–55, 2005.

[25] Whitney K. Newey and James L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–47, 1987.

[26] James L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155, 1986.

[27] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[28] Kenneth J. Rothman, Timothy L. Lash Associate Professor, and Sander Greenland. *Modern Epidemiology*. Wolters Kluwer, 3 edition edition, 2012-12-17.

[29] Sabine Schnabel and Paul Eilers. Optimal expectile smoothing. *Computational Statistics and Data Analysis*, 53:4168–4177, 2009.

[30] Fabian Sobotka, Göran Kauermann, Linda Schulze Waltrup, and Thomas Kneib. On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23(2):135–148, 2013.

[31] Fabian Sobotka and Thomas Kneib. Geoadditive expectile regression. *Computational Statistics and Data Analysis*, 2010. doi: 10.1016/j.csda.2010.11.015.

[32] Fabian Sobotka, Sabine Schnabel, Linda Schulze Waltrup, Paul Eilers, Thomas Kneib, and Goeran Kauermann. *expectreg: Expectile and Quantile Regression*, 2014. R package version 0.39.

[33] Linda Schulze Waltrup, Fabian Sobotka, Thomas Kneib, and Göran Kauermann. Expectile and quantile regression—david and goliath? *Statistical Modelling*, page 1471082X14561155, 2014-12-26.

[34] K. Yu, Z. Lu, and J. Stander. Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society Series D: The Statistician*, 52(3):331–350, 2003.