Journal of Econometrics 157 (2010) 396–408

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Robust penalized quantile regression estimation for panel data ☆

Carlos Lamarche *

Department of Economics, University of Oklahoma. 321 Hester Hall, 729 Elm Avenue, Norman, OK 73019, United States

ARTICLE INFO

Article history:
Received 30 October 2006
Received in revised form
3 February 2010
Accepted 12 March 2010
Available online 21 March 2010

Keywords:
Shrinkage
Robust
Quantile regression
Panel data
Individual effects

ABSTRACT

This paper investigates a class of penalized quantile regression estimators for panel data. The penalty serves to shrink a vector of individual specific effects toward a common value. The degree of this shrinkage is controlled by a tuning parameter λ. It is shown that the class of estimators is asymptotically unbiased and Gaussian, when the individual effects are drawn from a class of zero-median distribution functions. The tuning parameter, λ, can thus be selected to minimize estimated asymptotic variance. Monte Carlo evidence reveals that the estimator can significantly reduce the variability of the fixed-effect version of the estimator without introducing bias.

1. Introduction

Panel data consisting of multiple observations on individuals, firms, etc., over time provides an opportunity for analyzing economic relations while controlling for unobserved individual heterogeneity. However, classical least squares estimation methods designed for Gaussian models are often inadequate for empirical analysis. For example, Horowitz and Markatou (1996) find that the error term of an earning model, using CPS data, is not normally distributed. Moreover, the problems with least squares methods are highlighted in Angrist et al. (2002) longitudinal analysis of a voucher program. On the one hand, the within transformation gets rid of the time-invariant treatment indicator, and on the other hand, the random effects exclusive focus of the program's effect on the mean, is a limitation if the interest is on the lower quantiles.

Koenker (2004) suggested a quantile regression approach for panel data. He introduced a class of penalized quantile regression estimators providing a novel solution to the recognized difficulties of quantile regression for additive random effects models (Koenker and Hallock, 2000; Abrevaya and Dahl, 2008). The standard least squares transformations to deal with a large number of parameters are not available in quantile regression, so the approach proposes to estimate directly a vector of individual effects. The estimation of these (nuisance) parameters increases the variability of the estimates of the covariate effects, but regularization, or shrinkage of these effects toward a common value helps to reduce this additional variability. An ℓ_1 penalty term (Tibshirani, 1996; Donoho et al., 1998) serves to shrink the vector of individual effects, and a tuning parameter λ controls the degree of this shrinkage. Although Koenker (2004) shows that some degree of shrinkage is often desirable, finding precisely the value of λ remains unclear outside idealized Gaussian conditions. This paper investigates this issue showing that the class of penalized estimators for models with exogenous regressors is asymptotically unbiased and Gaussian. The parameter λ can thus be selected to minimize estimated asymptotic variance.

It has been acknowledged by others that the optimal choice of the regularization parameter λ is an interesting problem of both theoretical and practical importance (Zou et al., 2007). For instance, in model selection, the regularization parameter is selected by AIC (Akaike, 1973) and BIC (Schwartz, 1978), in ridge regression, it is selected by minimizing mean square error (Hoerl and Kennard, 1988), and in non-parametrics, the choice of λ is analogous to selecting the smoothing parameter by cross-validation (e.g. CV and GCV). In panel data, the classical random effects approach suggests maximum likelihood (MLE) or generalized least squares (GLS) methods, but preliminary Gaussian λ selection strategies under non-classical assumptions could lead to incorrect inference.

☆ This version: January 31, 2010. This paper is based on Chapter 1 of my dissertation "Quantile Regression for Panel Data" at the University of Illinois at Urbana-Champaign. I especially thank my advisor Roger Koenker for advice, insightful comments and suggestions. I am also grateful to Rick Ashley, Anil Bera, Bruce Hansen, Zhongjun Qu, Aris Spanos, and Zhijie Xiao for comments and valuable conversations, as well as to the seminar participants at UC Merced, University of Oklahoma, Virginia Tech, Robert Bohrer workshop at the University of Illinois at Urbana-Champaign, the 16th Midwest Econometric Group Conference, and the 14th International Conference on Panel Data. I am grateful to the Editor, Associate Editor, and four referees for their comments which have improved the paper considerably.

* Tel.: +1 405 325 5857.
E-mail address: lamarche@ou.edu.

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.03.042

The approach considered in this paper offers a robust alternative for λ selection.

This paper establishes the asymptotic theory of the penalized estimator deriving an asymptotic approximation for the optimal value of the tuning parameter. The main theoretical contribution is a decomposition of the penalty into two terms that depend on λ. The first term is asymptotically Gaussian, and the second term has a deterministic quadratic contribution to the limiting form of the objective function. The extension leads to an asymptotically unbiased estimator when the individual effects are drawn from a class of zero-median distribution functions. Because the asymptotic variance is a strictly convex function of the parameter, the optimal λ exists, is unique, and gives the minimum variance estimator in the class of penalized quantile regression estimators, the analog of the GLS in the class of penalized least squares estimators for panel data.

The next section presents the model and estimator. Section 3 is devoted to the asymptotic behavior of the estimator and to obtain the optimal tuning parameter. In Section 4, we offer Monte Carlo evidence. Section 5 provides conclusions.

2. Panel data models and methods

Consider the classical Gaussian random effects model

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T \tag{2.1}$$

where y_{it} is the dependent variable, $\boldsymbol{x}_{it} = (1, x_{it,2}, \ldots, x_{it,p-1})'$ is the vector of independent variables, the α_i's are unobservable time-invariant effects distributed independently across subjects, and u_{it} is the error term. It is convenient to write Eq. (2.1) as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{u}, \tag{2.2}$$

where the vector $\boldsymbol{y}$ is NT by 1, $\boldsymbol{X}$ is NT by p, $\boldsymbol{Z}$ is a NT by N "incidence matrix" of dummy variables, and $\boldsymbol{\alpha}$ and $\boldsymbol{u}$ are independent random vectors.

In this paper, we consider a conditional quantile model of the form,

$$Q_{Y_{it}}(\tau_j|\boldsymbol{x}_{it}, \alpha_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j) + \alpha_i \tag{2.3}$$

for all quantiles τ_j in the interval $(0, 1)$. The parameter $\boldsymbol{\beta}(\tau_j)$ models the covariate effect providing an opportunity for investigating how factors influence the location, scale and shape of the conditional distribution of the response. For instance, if we have an iid error term distributed as F and one covariate, the quantile functions $Q_{Y_{it}}(\tau_j|x_{it}, \alpha_i)$ are parallel lines with parameter $\boldsymbol{\beta}(\tau_j) = (\beta_0(\tau_j), \beta_1)' = (\beta_0 + F_u(\tau_j)^{-1}, \beta_1)'$. The model (2.3) also includes individual intercepts. We assume that the individual effect does not represent a distributional shift, since it is unrealistic to estimate it when the number of observations on each individual is small (Koenker, 2004). The individual specific effect is then a pure location shift effect α_i on the conditional quantiles of the response, implying that the conditional distribution for each subject have the same shape, but different locations as long as the α_i's are different.

Koenker's (2004) interpretation of the Gaussian random effects estimator as the penalized least squares estimator for the fixed effects extends the scope of quantile regression to panel data models. Koenker introduces a class of penalized estimators as the solution of,

$$\min_{\boldsymbol{\beta},\boldsymbol{\alpha}} \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N} \omega_j \rho_{\tau_j}(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) + \lambda \sum_{i=1}^{N} |\alpha_i| \tag{2.4}$$

where $\rho_{\tau_j}(u) = u(\tau_j - I(u \leq 0))$ is the quantile loss function, and ω_j is a relative weight given to the jth quantile. The weights control the influence of the quantiles on the estimation of the individual effects and the penalty $\lambda\|\boldsymbol{\alpha}\|_1$ serves to shrink the individual effects estimates toward zero to improve the performance of the estimate of $\boldsymbol{\beta}$. The ℓ_1 penalty function is convenient from the computational point of view since preserves the linear programming problem. The advantages of ℓ_1 shrinkage over ℓ_2 Gaussian shrinkage have been recognized in Tibshirani (1996) and Donoho et al. (1998). Note that in addition to estimating the covariate effects $\boldsymbol{\beta}(\tau)$'s, we estimate the α's using a penalty term that may reflect prior beliefs on the random effects (see, e.g., Koenker (2004), Geraci and Bottai (2007), Robinson (1991) and Ruppert et al. (2003), for quantile and least squares estimation).

The selection of λ represents a fundamental issue. Empirical researchers interested in applying quantile regression on panel data models need to arbitrarily chose the shrinkage parameter, affecting inference for the parameter of interest $\boldsymbol{\beta}(\boldsymbol{\tau})$. The solution of (2.4) is obviously sensitive to the value of λ, but in contrast to the least squares case that considers $\lambda = \sigma_u^2/\sigma_\alpha^2$ under Gaussian assumptions as illustrated in Koenker (2004, 2005), the literature lacks of a clear choice for $\lambda > 0$ (see, e.g., Koenker (2004), Geraci and Bottai (2007)).

This paper establishes the asymptotic theory of the penalized quantile regression estimator deriving an asymptotic approximation for the optimal value of λ. Under certain conditions, as will become clear in our later analysis, the estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ is asymptotically unbiased for all positive λ, therefore it is reasonable to consider choosing λ to minimize asymptotic variance. Our optimal choice of λ will give the minimum variance estimator in the class of penalized quantile regression estimators, the analog of the GLS in the class of penalized least squares estimators for panel data.

3. Asymptotic theory

We derive the asymptotic distribution of the penalized quantile regression estimator for several quantiles simultaneously estimated when the number of cross sectional units N and the number of time periods T both go to infinity. We use notation that is standard. The symbol "$\to$" signifies convergence in probability, "$\rightsquigarrow$" denotes convergence in distribution, sgn(a) is the sign of a, and tr A means trace of the matrix A.

The solution of (2.4) $(\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)', \hat{\boldsymbol{\alpha}}(\lambda)')'$ will be called the τ_j-th panel data quantiles. We write $\rho_{\tau_j}(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}(\tau_j) - \hat{\alpha}_i)$ as $\rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j) - \hat{\delta}_{0i}/\sqrt{T} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\delta}}_1(\tau_j)/\sqrt{TN})$, where $\xi_{it}(\tau_j) = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j)$ and,

$$\hat{\boldsymbol{\delta}}_0 = \begin{pmatrix} \hat{\delta}_{01} \\ \vdots \\ \hat{\delta}_{0N} \end{pmatrix} = \begin{pmatrix} \sqrt{T}(\hat{\alpha}_1 - \alpha_1) \\ \vdots \\ \sqrt{T}(\hat{\alpha}_N - \alpha_N) \end{pmatrix};$$

$$\hat{\boldsymbol{\delta}}_1 = \begin{pmatrix} \hat{\boldsymbol{\delta}}_1(\tau_1) \\ \vdots \\ \hat{\boldsymbol{\delta}}_1(\tau_J) \end{pmatrix} = \begin{pmatrix} \sqrt{TN}(\hat{\boldsymbol{\beta}}(\tau_1) - \boldsymbol{\beta}(\tau_1)) \\ \vdots \\ \sqrt{TN}(\hat{\boldsymbol{\beta}}(\tau_J) - \boldsymbol{\beta}(\tau_J)) \end{pmatrix}.$$

Moreover, we write $|\hat{\alpha}_i| = |\hat{\alpha}_i - (\alpha_i - \alpha_i)| = |\alpha_i + \hat{\delta}_{0i}/\sqrt{T}|$. For a given $\lambda = \lambda_T$, the minimization of (2.4) is equivalent to the following problem:

$$\begin{aligned}\min_{\delta_0,\delta_1} V_{TN}(\boldsymbol{\delta}) = \min_{\delta_0,\delta_1} \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N} \omega_j \Big\{ & \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j) - \delta_{0i}/\sqrt{T} \\ & - \boldsymbol{x}_{it}'\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}) - \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \Big\} \\ & + \lambda_T \left\{ \sum_{i=1}^{N} |\alpha_i + \delta_{0i}/\sqrt{T}| - |\alpha_i| \right\}.\end{aligned}$$

Assume the following regularity conditions,

A1. The variables y_{it} are independent with conditional (on $\boldsymbol{x}_{it}$, and α_i) distribution F_{it}, and continuous densities f_{it} uniformly bounded away from 0 and ∞, with bounded derivatives f'_{it} everywhere.

A2. The variables α_i, stochastically independent of $\boldsymbol{x}_{it}$ for all t, are exchangeable, identically, and independently distributed with zero median distribution G, and continuous densities g uniformly bounded away from 0 and ∞, with bounded derivatives g', for $i = 1, \ldots, N$.

A3. There exist positive definite matrices $\boldsymbol{H}_0$, $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, and $\boldsymbol{H}_3$ such that

$$\boldsymbol{H}_0 = \lim_{T,N\to\infty} \frac{1}{TN} \begin{pmatrix} \Omega_{11}\boldsymbol{X}'\boldsymbol{M}_1'\boldsymbol{M}_1\boldsymbol{X} & \ldots & \Omega_{1J}\boldsymbol{X}'\boldsymbol{M}_1'\boldsymbol{M}_J\boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \Omega_{1J}\boldsymbol{X}'\boldsymbol{M}_J'\boldsymbol{M}_1\boldsymbol{X} & \ldots & \Omega_{JJ}\boldsymbol{X}'\boldsymbol{M}_J'\boldsymbol{M}_J\boldsymbol{X} \end{pmatrix}$$

$$\boldsymbol{H}_1 = \lim_{T,N\to\infty} \frac{1}{TN} \begin{pmatrix} \omega_1\boldsymbol{X}'\boldsymbol{M}_1'\boldsymbol{\Phi}_1\boldsymbol{M}_1\boldsymbol{X} & \ldots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \ldots & \omega_J\boldsymbol{X}'\boldsymbol{M}_J'\boldsymbol{\Phi}_J\boldsymbol{M}_J\boldsymbol{X} \end{pmatrix}$$

$$\boldsymbol{H}_2 = \lim_{T,N\to\infty} \frac{1}{4TN} \begin{pmatrix} \tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1 & \ldots & \tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_J \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{X}}_J'\tilde{\boldsymbol{X}}_1 & \ldots & \tilde{\boldsymbol{X}}_J'\tilde{\boldsymbol{X}}_J \end{pmatrix}$$

$$\boldsymbol{H}_3 = \lim_{T,N\to\infty} \frac{1}{TN} \begin{pmatrix} \tilde{\boldsymbol{X}}_1'\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_1 & \ldots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \ldots & \tilde{\boldsymbol{X}}_J'\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_J \end{pmatrix}$$

where $\Omega_{kl} = \omega_k(\tau_k \wedge \tau_l - \tau_k\tau_l)\omega_l$, $\boldsymbol{M}_j = \boldsymbol{I} - \boldsymbol{P}_j$, $\boldsymbol{P}_j = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{\Phi}_j\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}_j$, $\boldsymbol{\Phi}_j = \mathrm{diag}(f_{it}(\xi_{it}(\tau_j)))$, $\boldsymbol{\Psi} = \mathrm{diag}(g(0))$, and $\tilde{\boldsymbol{X}}_j = (\boldsymbol{Z}'\boldsymbol{\Phi}_j\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}_j\boldsymbol{X}$.

A4. $\max \|\boldsymbol{x}_{it}\|/\sqrt{TN} \to 0$.

A5. There exists a constant $c > 0$ such that $N^c/T \to 0$.

A6. The regularization parameter $\lambda_T/\sqrt{T} \to \lambda \geq 0$.

Koenker (2004) considers similar conditions. The behavior of the conditional density in a neighborhood of $\xi_{it}(\tau)$ is crucial for the asymptotic behavior of the quantile regression estimator. Condition A1 ensures a well-defined asymptotic behavior of the quantile regression estimator. The independence on the y_{it}'s, conditional on α_i, is assumed in Koenker (2004) and Geraci and Bottai (2007). The zero median assumption in condition A2 is an additional condition to the ones assumed in Koenker (2004), and it could be interpreted as a random-effects type of assumption. Condition A2 suggests a Knight's (1998) decomposition of the penalty term and also plays a crucial role ensuring a well-defined asymptotic behavior of the estimator. In condition A3, the existence of the limiting form of the positive definite matrices is used to invoke the Lindeberg–Feller Central Limit Theorem. In particular, while $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$, are used in the piece of objective function that corresponds to the standard quantile regression problem, $\boldsymbol{H}_2$ and $\boldsymbol{H}_3$ are used in the piece of objective function that corresponds to the penalty term. Condition A4 is important both for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function. Condition A5 is needed to make sure that the contribution of the remainder term that comes from the Bahadur representation of the individual effects is asymptotically negligible. Condition A6 is required to achieve a square root-n consistency.

Theorem 1. *Under the regularity conditions A1–A6, the minimizer of the objective function*

argmin $V_{TN}(\boldsymbol{\delta}_1) \rightsquigarrow$ *argmin* $V_0(\boldsymbol{\delta}_1)$,

where $V_0(\boldsymbol{\delta}_1) = -\boldsymbol{\delta}_1'(\boldsymbol{B} + \lambda\boldsymbol{C}) + \frac{1}{2}\boldsymbol{\delta}_1'(\boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3)\boldsymbol{\delta}_1$, *and* $\boldsymbol{B}$, *and* $\boldsymbol{C}$ *are zero mean Gaussian independent vectors with covariance matrices* $\boldsymbol{H}_0$, *and* $\boldsymbol{H}_2$, *respectively.*

Remark 1. The Bahadur representation of the unobserved specific effects contains only one "interesting" term, which depends on $\boldsymbol{\delta}_1(\tau)$. Lemma 1 shows that the asymptotic contributions of the terms that do not depend on the parameter of interest are negligible confirming that the finite dimensional convergence holds, without the necessity of cumbersome algebra.

Remark 2. The contribution of the penalty to the limiting objective function leads to a variation of the existing quantile regression asymptotic theory. Theorem 2's limiting objective function $V_0(\cdot)$ in Koenker (2004) can be written as,

$$-\boldsymbol{\delta}_1'(\boldsymbol{B} + \lambda\boldsymbol{s}) + \frac{1}{2}\boldsymbol{\delta}_1'\boldsymbol{H}_1\boldsymbol{\delta}_1$$

for $\boldsymbol{s} = (\mathrm{sgn}(\alpha_i))$. This leads to an asymptotically normal estimator with mean $\boldsymbol{\beta}(\boldsymbol{\tau}) + \lambda\mathbb{E}\boldsymbol{s}$ and variance $\boldsymbol{H}_1^{-1}\boldsymbol{H}_0\boldsymbol{H}_1^{-1}$. Notice that the selection of the tuning parameter in this setting suggests to set λ equal to zero because the variance does not depend on λ. Condition A2 suggests a Knight's (1998) decomposition on the penalty term, leading to an asymptotically unbiased estimator whose variance depends on λ, as shown in Corollary 1.

Corollary 1. *Let* $\boldsymbol{\Gamma}_0(\lambda) = \boldsymbol{H}_0 + \lambda^2\boldsymbol{H}_2$, *and* $\boldsymbol{\Gamma}_1(\lambda) = \boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3$. *Under the conditions of Theorem 1, the penalized estimator* $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ *is asymptotically normally distributed with mean* $\boldsymbol{\beta}(\boldsymbol{\tau})$ *and covariance matrix* $\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}$.

The estimation of one quantile $\tau_j = \tau$ can be seen as a special case with the number of quantiles J and the weight ω set to unity. It is convenient to define,

$$\boldsymbol{D}_0 = \lim_{T,N\to\infty} \frac{1}{TN}\tau(1-\tau)\boldsymbol{X}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{X}; \qquad \boldsymbol{D}_2 = \lim_{T,N\to\infty} \frac{1}{4TN}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}},$$

$$\boldsymbol{D}_1 = \lim_{T,N\to\infty} \frac{1}{TN}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{\Phi}\boldsymbol{M}\boldsymbol{X}; \qquad \boldsymbol{D}_3 = \lim_{T,N\to\infty} \frac{1}{TN}\tilde{\boldsymbol{X}}'\boldsymbol{\Psi}\tilde{\boldsymbol{X}}.$$

Corollary 2. *Let* $\boldsymbol{\Sigma}_0(\lambda) = \boldsymbol{D}_0 + \lambda^2\boldsymbol{D}_2$, *and* $\boldsymbol{\Sigma}_1(\lambda) = \boldsymbol{D}_1 + 2\lambda\boldsymbol{D}_3$. *Under the conditions of Theorem 1, the penalized estimator* $\hat{\boldsymbol{\beta}}(\tau, \lambda)$ *is asymptotically normally distributed with mean* $\boldsymbol{\beta}(\tau)$ *and covariance matrix* $\boldsymbol{\Sigma}_1(\lambda)^{-1}\boldsymbol{\Sigma}_0(\lambda)\boldsymbol{\Sigma}_1(\lambda)^{-1}$.

Additionally, the fixed effects estimator is a special case with λ equal to zero. In the iid case $\{u_{it}\} \sim F$, we obtain an estimator that is asymptotically similar to the classical fixed effects estimator,

$$\sqrt{TN}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}\left(\boldsymbol{0}, \omega^2\boldsymbol{D}_0^{-1}\right),$$

where $\omega^2 = \tau(1-\tau)/f^2$, $f_{it}(\xi_{it}(\tau)) = f$ for all i, t, and $\boldsymbol{D}_0 = \lim_{N,T\to\infty} \frac{1}{TN}\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}$. The asymptotic relative efficiency (ARE) of the fixed effects quantile regression estimator $\hat{\beta}_j(\tau)$ relative to the classical fixed effects estimator, which is simply the ratio of the asymptotic variances, in the iid case yields ARE $= \omega^2/\sigma_u^2$. Therefore, the fixed effects case gives the standard result that the median quantile regression estimator has smaller asymptotic variance than the least squares estimator if $(2f)^{-1} < \sigma_u$.

Corollary 3. *Under the conditions of Theorem 1, the penalized quantile regression estimator* $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda^*)$ *is asymptotically normally distributed with mean* $\boldsymbol{\beta}(\boldsymbol{\tau})$ *and covariance matrix* $\boldsymbol{\Gamma}_1(\lambda^*)^{-1}\boldsymbol{\Gamma}_0(\lambda^*)\boldsymbol{\Gamma}_1(\lambda^*)^{-1}$ *where* $\lambda^* = \arg\min\{\mathrm{tr}\,\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\}$.

Recall that the penalized estimator is asymptotically unbiased for all values of λ, but the shrinkage of the individual effects controlled by the tuning parameter affects the asymptotic variance. Corollary 3 suggests choosing λ to minimize asymptotic variance, which under the conditions of Theorem 1 implies minimizing asymptotic MSE. It is therefore of interest to investigate if we can in fact minimize variance by setting the tuning parameter $\lambda = \lambda^*$.

The primary objective is now to show that the optimal tuning parameter λ^* exists and it is unique under the regularity conditions. The result, formally established in Theorem 2, holds under the conditions of Theorem 1 and Lemmas 2–5. Lemmas 2 and 3 show that there exists a non-empty and compact set, defined by the eigenvalues ζ of the asymptotic covariance matrix. Lemmas 4 and 5 show that the rational functions, obtained after we derive the spectral decomposition of the asymptotic covariance matrix, are strictly convex.

Theorem 2. *Under conditions A1–A6, there exists a unique minimizer*

$$\lambda^* = \arg\min_{\lambda\in\mathcal{D}} \left\{ \operatorname{tr} \boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1} \right\}.$$

Theorem 2 shows the existence and uniqueness of λ^* and suggests that it is possible to have an estimator that is, on average, more efficient than the fixed effects and pooled quantile regression estimators. Because all the diagonal elements of $\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}$ are equally weighted, the optimal shrinkage parameter λ^* may be influenced by scale effects.[1] For instance, a large $\mathrm{Var}\left(\hat{\beta}_k(\tau_j, \lambda)\right)$ may have a distorting effect on $\sum_j \sum_k \mathrm{Var}\left(\hat{\beta}_k(\tau_j, \lambda)\right)$ if the shape of the variances are not similar. One possibility out of this issue is to normalize the covariance matrix. Another possibility, which is considered in this paper, is to select one element of the main diagonal of the asymptotic covariance matrix $\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}$. The following corollary derives the exact amount of shrinkage that leads to the minimum variance quantile regression estimator for a single parameter, say β_k.

A7. The k-th column of $\boldsymbol{X}$ is a time-varying covariate $\boldsymbol{x}_k = (x_{11,k}, \ldots, x_{NT,k})'$ such that $x_{it,k} \neq x_{is,k}$ for at least for one s.

Corollary 4. *Under conditions A1–A7, the unique optimal value of the regularization parameter that minimizes the asymptotic variance of the penalized quantile regression estimator $\hat{\beta}_k(\tau_j, \lambda^*)$ is,*

$$\lambda^* = \frac{8\Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk}}{\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk}}.$$

Remark 3. In the iid case for the median the optimal tuning parameter is simply $\lambda^* = g/f$, which shrinks in the same way than $\lambda = \sigma_u/\sigma_\alpha$. When the scale parameter of the distribution of the α_i tends to infinity, we see that the optimal λ tends to zero. In contrast, when the scale parameter tends to zero, the optimal λ tends to infinity.

Remark 4. Because we do not need to assume (i) a Gaussian structure, (ii) existence of second moment (e.g., α_i or u_{it} can be assumed to be drawn from a Cauchy distribution) and (iii) spherical errors, the optimal λ^* appears as a robust alternative to $\lambda = \sigma_u/\sigma_\alpha$. Although we will be forced to estimate the unknown parameter λ^*, there will be no need to guarantee non-negative estimates of σ_α, a frequent problem well documented in the literature.

[1] I am grateful to Professor B. Hansen for pointing out this.

3.1. Estimation

This section briefly suggests how to construct an estimator, $\hat{\lambda}$, for the optimal degree of shrinkage, λ^*, that crucially depends on estimation of the conditional density $f(\xi(\tau))$ and the density of the individual effects $g(0)$. The estimation of $f(\xi(\tau))$ in iid and non-iid settings require the use of standard quantile regression techniques (see, e.g., Section 3.4 in Koenker (2005) for a review of methods). The estimation of the conditional quantiles $\xi(\tau, \lambda)$ requires residuals $\hat{u}(\tau, \lambda)$ evaluated at $\lambda = 0$. The estimation of $g(0)$ is based on individual effects $\hat{\alpha}_i(0)$, denoted below simply as $\hat{\alpha}_i$. If this idea could give us a consistent estimate of g, the first asymptotic requirement should be that $\hat{\alpha}_i \to \alpha_i$. Under the conditions of Koenker's (2004) Theorem 1, and previous regularity conditions, it can be shown that the fixed effects estimator $\hat{\alpha}_i \to \alpha_i$. The second minimal requirement should be $\hat{g} \to g$. We consider consistent estimation strategies including kernel methods, $(1/Nh_N)\sum_i K(\hat{\alpha}_i/h_N)$, on a sample of individual fixed effects estimates $\{\hat{\alpha}_1, \ldots, \hat{\alpha}_N\}$, using a Gaussian kernel $K(\cdot)$ and a bandwidth h_N.

Inference can be considered by accommodating standard quantile regression tests. Wald-type statistics (see, e.g., Koenker and Bassett, 1982) can be considered for basic general linear hypothesis on a vector $\boldsymbol{\xi}$ of the form $\mathrm{H}_0 : \boldsymbol{R}\boldsymbol{\xi} = \boldsymbol{r}$, where $\boldsymbol{R}$ is a matrix that depends on the type of restrictions imposed. We could evaluate the significance of differences across coefficient estimates of a vector $\boldsymbol{\xi} = (\beta_1(\tau, \lambda^*), \ldots, \beta_{p-1}(\tau, \lambda^*))'$. More general hypothesis including evaluating the vector over a range of quantiles could be accommodated by considering the Kolmogorov–Smirnov statistics (see, e.g., Koenker and Xiao (2002), and Koenker (2005)).

4. Monte Carlo results

In this section we will report the results of several simulation experiments designed to evaluate the finite sample performance of the method considered above. First, we will investigate the bias and variance of the penalized estimator. Second, we will study the performance of $\hat{\lambda}$ as an estimator of λ^*, and then we will contrast the performance of the penalized quantile regression estimator with classical panel data estimators.

Two versions of the model are considered in the simulation experiments. When the dependent variable is generated from the location shift model,

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it}, \tag{4.1}$$

the corresponding quantile regression model is,

$$Q_{Y_{it}}(\tau_j|x_{it}, \alpha_i) = \beta_0(\tau_j) + \beta_1 x_{it} + \alpha_i,$$

where $\beta_0(\tau_j) = \beta_0 + F_u(\tau_j)^{-1}$. Note that this is a particular case of model (2.3) where $\beta_1(\tau_j) = \beta_1$ for all $\tau_j \in (0, 1)$. In the case of the location-scale shift model,

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \gamma_0(1 + \gamma_1 x_{it})u_{it}, \tag{4.2}$$

the corresponding quantile regression model is,

$$Q_{Y_{it}}(\tau_j|x_{it}, \alpha_i) = \beta_0(\tau_j) + \beta_1(\tau_j)x_{it} + \alpha_i$$

where $\beta_0(\tau_j) = \beta_0 + \gamma_0 F_u(\tau_j)^{-1}$ and $\beta_1(\tau_j) = \beta_1 + \gamma_0\gamma_1 F_u(\tau_j)^{-1}$. We consider $\alpha_i \sim G$ and $u_{it} \sim F$. The independent variable x_{it} is generated as in Koenker (2004) and the parameters of the model are assumed to be 0 and 1 for the intercept and slope, respectively.

Considering the location-shift model and assuming that α_i and u_{it} are iid $\mathcal{N}(0, 1)$, we evaluate the small sample performance of λ^* generating data from a small data set: $N = 50$ and $T = 5$. Because α_i and u_{it} are independent and identically distributed

Table 4.1
Bias and RMSE of $\hat{\lambda}$ for the optimal amount of shrinkage λ^*.

		Bias				RMSE			
		Log	Kernel Methods			Log	Kernel Methods		
		Spline	S86	S92	UCV	Spline	S86	S92	UCV
N	T	Quantile $\tau = 0.25$ — location shift model							
50	5	0.1207	0.2532	0.2701	0.2729	0.3506	0.4225	0.4291	0.4566
50	15	0.0895	0.0887	0.1100	0.1129	0.3785	0.2677	0.2752	0.3144
50	25	0.0406	0.1152	0.1346	0.1391	0.3145	0.2861	0.2923	0.3155
100	5	0.1356	0.2454	0.2593	0.2596	0.3276	0.3659	0.3723	0.3949
100	15	0.0792	0.0677	0.0864	0.0851	0.3145	0.2124	0.2213	0.2455
100	25	0.0277	0.0993	0.1177	0.1171	0.2399	0.2261	0.2334	0.2631
500	5	0.1967	0.2390	0.2480	0.2441	0.2729	0.2987	0.3036	0.3099
500	15	0.0060	0.0499	0.0606	0.0566	0.1106	0.1276	0.1330	0.1404
500	25	0.0343	0.0838	0.0933	0.0928	0.1266	0.1499	0.1546	0.1611
N	T	Quantile $\tau = 0.50$ — location shift model							
50	5	0.0586	0.1965	0.2149	0.2178	0.3229	0.3810	0.3869	0.4125
50	15	0.0735	0.0973	0.1194	0.1209	0.3670	0.2780	0.2848	0.3382
50	25	0.0738	0.0916	0.1103	0.1180	0.3321	0.2534	0.2606	0.2941
100	5	0.0620	0.1796	0.1947	0.2003	0.2796	0.3135	0.3192	0.3423
100	15	0.0275	0.1049	0.1214	0.1269	0.2431	0.2321	0.2394	0.2800
100	25	0.0511	0.0862	0.1027	0.1062	0.2511	0.2073	0.2134	0.2441
500	5	0.0984	0.1508	0.1610	0.1569	0.1861	0.2162	0.2217	0.2280
500	15	0.0213	0.0786	0.0897	0.0836	0.1089	0.1474	0.1524	0.1665
500	25	0.0030	0.0591	0.0692	0.0668	0.0989	0.1285	0.1332	0.1414
N	T	Quantile $\tau = 0.25$ — location-scale shift model							
50	5	0.0424	0.1192	0.1402	0.1460	0.3070	0.2896	0.2989	0.3299
50	15	0.1073	0.0553	0.0753	0.0816	0.3880	0.2184	0.2226	0.3010
50	25	0.0685	0.0975	0.1168	0.1258	0.3420	0.2564	0.2617	0.3167
100	5	0.0249	0.1184	0.1350	0.1323	0.2341	0.2540	0.2618	0.2982
100	15	0.1241	0.0250	0.0438	0.0419	0.3396	0.1529	0.1612	0.1847
100	25	0.0784	0.0821	0.1010	0.0976	0.2846	0.2020	0.2097	0.2323
500	5	0.0007	0.0498	0.0608	0.0562	0.1114	0.1360	0.1417	0.1521
500	15	0.0728	0.0088	0.0024	0.0012	0.1762	0.0810	0.0692	0.0773
500	25	0.0079	0.0682	0.0784	0.0732	0.1005	0.1338	0.1382	0.1519
N	T	Quantile $\tau = 0.50$ — location-scale shift model							
50	5	0.1578	0.0130	0.0349	0.0450	0.4739	0.2118	0.2192	0.2774
50	15	0.1439	0.0252	0.0457	0.0557	0.4558	0.2021	0.2062	0.2435
50	25	0.1510	0.0190	0.0403	0.0474	0.4307	0.1754	0.1824	0.2420
100	5	0.1207	0.0278	0.0455	0.0453	0.3517	0.1723	0.1805	0.2014
100	15	0.1369	0.0130	0.0316	0.0355	0.3587	0.1484	0.1544	0.1894
100	25	0.1125	0.0312	0.0489	0.0498	0.3570	0.1596	0.1653	0.2022
500	5	0.0677	0.0107	0.0014	0.0069	0.1793	0.0929	0.0785	0.1029
500	15	0.0168	0.0419	0.0531	0.0496	0.1063	0.1066	0.1126	0.1234
500	25	0.0044	0.0571	0.0677	0.0597	0.0970	0.1251	0.1291	0.1403

Gaussian random variables with scale parameter equal to 1, a priori we expect that a tuning parameter equal to 1 would perform well.

Monte Carlo evidence, presented in Fig. 4.1, suggests that the estimator of the slope parameter has essentially zero bias for all λ, and its variance does change with λ. The variability of the estimator decreases first, then increases, and lastly it is constant in λ giving rise to the questions: What is the optimal λ parameter, and how can it be determined in this case? Letting $j = k = 2$ and $\tau_2 = \omega_2 = 0.5$, Corollary 4 gives $\lambda^* = 1$. Therefore, the method correctly chooses the value of the shrinkage parameter, reducing the variability of the fixed effect estimator 16% without sacrificing bias.

In the lower panels, we report the performance of the first five individual specific effects estimates. We see now that both bias and variance of $\hat{\alpha}_i$'s changes with λ. Individual effects are estimated using a small number of observations T, therefore small biases were expected. The bias vanishes when λ increases because, although α_i is poorly estimated, λ forces the individual effect estimator to be equal to the location parameter of the distribution of the individual effects. The lower-right panel shows that the variance of the shrunken coefficient goes to zero as λ increases.

Using Table 4.1, we now investigate the performance of the estimator $\hat{\lambda}$ in several models considering three quantiles $\tau = \{0.25, 0.50, 0.75\}$, sample sizes $N = \{50, 100, 500\}$ and $T = \{5, 15, 25\}$, and using the density estimation methods described above. The Monte Carlo results at the 0.25 quantile were similar to the results at the 0.75 quantile, so we will only report estimates at the 0.25 quantile.

We estimate λ^* assuming that the distribution function of the error term F and the distribution function of the individual specific effects G are Gaussian and unknown. Without loss of generality, we assume that the parameters of the model are equal to zero. While we estimate the scalar sparsity parameter in the location shift model, we use Powell's kernel method (implemented in Koenker, 2006) in the location-scale shift model. On the other hand, we use a logspline method (see, e.g., Kooperberg and Stone, 1991) and a Gaussian kernel with three different bandwidths to estimate g: Silverman's (1986) rule of thumb (S86), Scott's (1992) variation (S92), and Scott's (1992) unbiased cross validation (UCV).

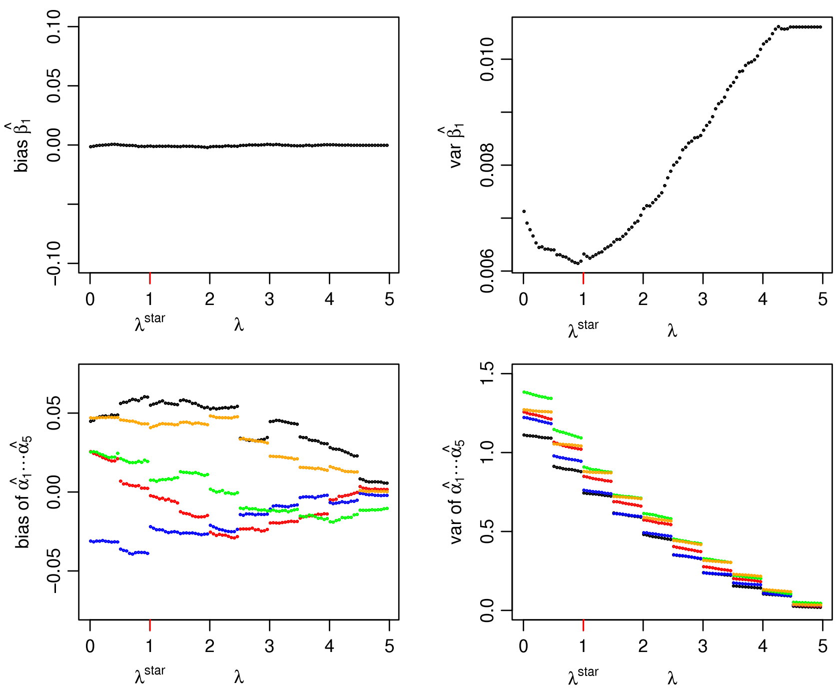

Fig. 4.1. Small sample performance of the quantile estimators as a function of the regularization parameter. The upper panels present the performance of the slope estimate, and the lower panels the performance of the first five individual effects estimates. Each dot represents a statistic based on 400 randomly generated samples. The additional line on the x-axis indicates λ^*.

We see that while the logspline method performs better than kernel methods in the location shift model, it performs worse than kernel methods in the location-scale shift model. We also observe that the bias has a tendency to decrease as we increase the sample size. For example, using the logspline method to estimate the density of the individual effects g in the location shift model, the bias is reduced from 5.9% ($N = 50$ and $T = 5$) to 0.3% ($N = 500$ and $T = 25$) at the 0.5 quantile. We also observe in Table 4.2 that as the sample size increases and consequently $\hat{\lambda}$ goes to λ^*, the feasible penalized estimator $\hat{\beta}_1(\tau_j, \hat{\lambda})$ is almost indistinguishable from the unfeasible penalized estimator $\hat{\beta}_1(\tau_j, \lambda^*)$.[2]

Table 4.3 provides evidence on our previous Remark 4. We generate data from both the location-shift model and the location shift model presented in Eqs. (4.1) and (4.2) considering the individual effects α and the error term u distributed as Cauchy $\mathcal{C}$ as well as Gaussian $\mathcal{N}(0, 1)$, t-student with three degrees of freedom t_3, and a mixture distribution defined as $0.95\mathcal{N}(0, 1) + 0.05\mathcal{C}$. The table shows the performance of three λ^* selection strategies: (1) Laplacian strategies $\tilde{\lambda} = \hat{\sigma}_u/\hat{\sigma}_\alpha$ (2) Gaussian strategies considering $\check{\lambda} = \hat{\sigma}_u^2/\hat{\sigma}_\alpha^2$, and (3) the robust method proposed in this paper, $\hat{\lambda}$. As expected, both the Laplacian strategy $\tilde{\lambda}$ and the robust optimal strategy $\hat{\lambda}$ offer similar performances in the classical Gaussian case, but $\hat{\lambda}$ provides a superior performance across all the variants of the models and distributions.

We compare, in Table 4.4, the performance outside the class of quantile regression estimators considering the following estimators: (1) the ordinary least squares (OLS); (2) the fixed effects estimator (FE); (3) the GLS estimator; (4) the pooled quantile regression estimator (QR); (5) the fixed effect quantile regression estimator (FEQR); (6) the penalized quantile regression (PQR).

We expand the design of the experiment. The y variables are generated as before, and the x series are generated using a similar method to that of Nerlove (1971) and Baltagi (1981), having an intraclass correlation coefficient $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_u^2)$ either $\{0.4, 0.8\}$ for a fixed number of cross sectional and time data points $N = 50$ and $T = 5$, a small panel data set representing an unfavorable scenario for λ selection. The parameters of the models are assumed to be 0 and 1 for the intercept and the slope, respectively.[3] We report the results based on 400 replications.

Motivated by Horowitz and Markatou (1996) finding of student t tail behavior, we investigate the relative performance outside the Gaussian setting considering three alternatives to the standard case of error components normally distributed ($\mathcal{N}$). The variables are drawn from a t-student distribution with 3 degrees of freedom (t_3), and also from different distributions (e.g., α_i is drawn from $\mathcal{N}$ and u_{it} from t_3).[4]

In Table 4.4, we see that the PQR is unbiased, robust, and it attains the minimum variance in the class of quantile regression estimators for longitudinal data. The maximum bias is 0.1%, and the estimator is more efficient than the true GLS estimator when $u_{it} \sim t_3$. When the intraclass correlation coefficient is positive, the penalized quantile regression estimator is considerably more efficient than both QR and FEQR. For instance, PQR offers relative to QR an average variance reduction of 39%, ranging from 6% to 75%.

[2] Geraci and Bottai (2007) propose a quasi-maximum likelihood estimator based on the asymmetric Laplace distribution, which has the advantage of eliminating the selection of λ. Using several simulation experiments, we compared the performance of their method with Koenker's shrinkage estimator that uses $\hat{\lambda}$. The evidence suggests that the feasible version of the penalized estimator offers a superior finite sample performance and considerably lower computational costs.

[3] The location-scale case imposes different considerations for the model and the estimators. The intraclass correlation coefficient is now $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \bar{\sigma}^2)$, where $\bar{\sigma}^2$ denotes the expected value of the conditional variance $\gamma_0^2(1 + \gamma_1 x)^2\sigma_u^2$. The parameter γ_0 is set equal to 1 when either one or two of the error components has Gaussian distribution. When $\alpha, u \sim t_3$, the variances of the random components are given, thus we use γ_0 to produce different values of the intraclass correlation coefficient. The parameter γ_0 takes the value 0.46 when $\rho = 0.4$, and the value 0.08 when $\rho = 0.8$. The parameter γ_1 is equal to 0.1 in all the variants of the model.

[4] When α_i and u_{it} are drawn from different distributions, the variance of the Gaussian variable is changed to give the value of ρ assumed in the experiment.

Table 4.2
Bias and root mean square error for the slope. PQR stands for the unfeasible penalized estimator $\hat{\beta}_1(\tau, \lambda^*)$ and FPQR denotes the Feasible Penalized Estimator $\hat{\beta}_1(\tau, \hat{\lambda})$.

		Bias				RMSE			
		QR	QRFE	PQR	FPQR	QR	FEQR	PQR	FPQR
N	T	Quantile $\tau = 0.25$ — location shift model							
50	5	0.0015	0.0062	0.0049	0.0048	0.0765	0.0793	0.0649	0.0665
50	15	0.0005	0.0005	0.0006	0.0009	0.0636	0.0434	0.0412	0.0410
50	25	0.0005	0.0016	0.0013	0.0012	0.0604	0.0323	0.0304	0.0303
100	5	0.0037	0.0056	0.0044	0.0039	0.0495	0.0575	0.0435	0.0441
100	15	0.0035	0.0014	0.0010	0.0012	0.0451	0.0307	0.0281	0.0278
100	25	0.0009	0.0001	0.0004	0.0005	0.0383	0.0231	0.0216	0.0215
500	5	0.0002	0.0005	0.0005	0.0005	0.0229	0.0276	0.0206	0.0212
500	15	0.0002	0.0002	0.0003	0.0003	0.0190	0.0140	0.0126	0.0126
500	25	0.0004	0.0002	0.0002	0.0002	0.0183	0.0110	0.0103	0.0103
N	T	Quantile $\tau = 0.50$ — location shift model							
50	5	0.0011	0.0025	0.0008	0.0001	0.0709	0.0745	0.0631	0.0626
50	15	0.0004	0.0015	0.0015	0.0017	0.0594	0.0420	0.0385	0.0392
50	25	0.0005	0.0022	0.0017	0.0017	0.0561	0.0306	0.0290	0.0289
100	5	0.0037	0.0068	0.0063	0.0058	0.0443	0.0552	0.0421	0.0415
100	15	0.0037	0.0029	0.0028	0.0028	0.0414	0.0308	0.0276	0.0275
100	25	0.0019	0.0003	0.0001	0.0000	0.0364	0.0228	0.0210	0.0210
500	5	0.0001	0.0008	0.0002	0.0001	0.0206	0.0268	0.0190	0.0189
500	15	0.0002	0.0005	0.0005	0.0005	0.0180	0.0136	0.0125	0.0125
500	25	0.0008	0.0000	0.0000	0.0000	0.0173	0.0109	0.0100	0.0100
N	T	Quantile $\tau = 0.25$ — location-scale shift model							
50	5	0.0208	0.0173	0.0136	0.0140	0.0796	0.0812	0.0663	0.0656
50	15	0.0172	0.0005	0.0014	0.0014	0.0653	0.0427	0.0385	0.0383
50	25	0.0232	0.0045	0.0050	0.0049	0.0605	0.0289	0.0275	0.0275
100	5	0.0256	0.0127	0.0121	0.0117	0.0511	0.0513	0.0406	0.0411
100	15	0.0213	0.0073	0.0061	0.0059	0.0444	0.0273	0.0254	0.0254
100	25	0.0225	0.0027	0.0026	0.0025	0.0448	0.0229	0.0216	0.0217
500	5	0.0245	0.0144	0.0123	0.0122	0.0235	0.0251	0.0202	0.0201
500	15	0.0229	0.0045	0.0044	0.0044	0.0191	0.0131	0.0119	0.0119
500	25	0.0194	0.0028	0.0026	0.0026	0.0190	0.0090	0.0085	0.0085
N	T	Quantile $\tau = 0.50$ — location-scale shift model							
50	5	0.0020	0.0041	0.0022	0.0022	0.0721	0.0794	0.0587	0.0597
50	15	0.0020	0.0033	0.0017	0.0016	0.0617	0.0422	0.0378	0.0382
50	25	0.0032	0.0018	0.0020	0.0023	0.0572	0.0288	0.0276	0.0278
100	5	0.0015	0.0032	0.0003	0.0005	0.0481	0.0501	0.0410	0.0407
100	15	0.0002	0.0024	0.0007	0.0006	0.0400	0.0257	0.0253	0.0259
100	25	0.0001	0.0000	0.0001	0.0001	0.0433	0.0225	0.0210	0.0210
500	5	0.0012	0.0003	0.0010	0.0010	0.0227	0.0245	0.0192	0.0190
500	15	0.0012	0.0001	0.0002	0.0002	0.0183	0.0126	0.0117	0.0116
500	25	0.0018	0.0004	0.0001	0.0001	0.0181	0.0089	0.0087	0.0087

Table 4.3
Comparison of empirical λ selection strategies for the penalized quantile regression estimator. The table shows averages of estimates relative to the optimal level of shrinkage λ^*, and .95% confidence intervals. $N = 50$ and $T = 5$. The dagger † indicates an entry greater than 10^3.

λ Selection	α and u distributions			
Strategy	Gaussian	t_3	Cauchy	Mixture
	Location-shift model			
Laplacian ($\tilde{\lambda}$)	1.041 [0.732, 1.350]	1.148 [0.334, 1.961]	† [−†, †]	† [−†, †]
Gaussian ($\check{\lambda}$)	1.108 [0.418, 1.798]	1.489 [−2.538, 5.516]	† [−†, †]	† [−†, †]
Optimal Shrinkage ($\hat{\lambda}$)	1.017 [0.401, 1.634]	0.936 [0.529, 1.343]	1.282 [0.623, 1.943]	1.019 [0.411, 1.626]
	Location-scale shift model			
Laplacian ($\tilde{\lambda}$)	1.304 [0.738, 1.870]	13.077 [−379.2, 405.4]	† [−†, †]	241.9 [−†, †]
Gaussian ($\check{\lambda}$)	2.408 [−0.306, 4.877]	† [−†, †]	† [−†, †]	† [−†, †]
Optimal Shrinkage ($\hat{\lambda}$)	0.934 [0.544, 1.325]	1.031 [0.583, 1.479]	1.363 [0.642, 2.085]	0.965 [0.577, 1.353]

Table 4.4
Small sample performance of panel data slope estimators. RE is a measure of relative efficiency defined as the standard error of the QR estimator relative to other panel data estimators, including PQR (e.g., $\mathrm{se}(\hat{\beta}_{1,PQR})/\mathrm{se}(\hat{\beta}_{1,QR})$).

ρ	λ^*	α	u	Statistic	Panel Data Estimators					
					Least Squares			Quantile Regression		
					OLS	FE	GLS	QR	FEQR	PQR
					Quantile $\tau = 0.5$ — location shift model					
0.0	5.00	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0010	0.0003	0.0010	0.0001	0.0021	0.0001
				RE	0.7711	2.0358	0.7711	1.0000	2.2510	1.0000
	5.00	t_3	t_3	Bias	0.0011	0.0042	0.0011	0.0002	0.0005	0.0002
				RE	1.2498	3.3005	1.2498	1.0000	2.5361	1.0000
	5.00	t_3	$\mathcal{N}$	Bias	0.0001	0.0010	0.0001	0.0008	0.0014	0.0008
				RE	0.7963	2.0503	0.7963	1.0000	2.2799	1.0000
	5.00	$\mathcal{N}$	t_3	Bias	0.0005	0.0042	0.0005	0.0001	0.0010	0.0001
				RE	1.2457	3.3514	1.2457	1.0000	2.5310	1.0000
0.4	1.35	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0001	0.0011	0.0003	0.0005	0.0002	0.0001
				RE	0.8816	1.2299	0.7956	1.0000	1.3095	0.9476
	1.25	t_3	t_3	Bias	0.0015	0.0009	0.0007	0.0005	0.0026	0.0004
				RE	1.4226	1.7116	1.2080	1.0000	1.2919	0.8860
	0.80	t_3	$\mathcal{N}$	Bias	0.0006	0.0013	0.0008	0.0015	0.0027	0.0007
				RE	1.2474	1.0138	0.9489	1.0000	1.1315	0.8521
	1.80	$\mathcal{N}$	t_3	Bias	0.0002	0.0023	0.0006	0.0003	0.0002	0.0001
				RE	1.0511	1.9527	1.0723	1.0000	1.5585	0.9694
0.8	0.30	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0007	0.0002	0.0003	0.0008	0.0002	0.0005
				RE	0.8871	0.5590	0.4975	1.0000	0.6450	0.5849
	0.35	t_3	t_3	Bias	0.0002	0.0013	0.0009	0.0004	0.0003	0.0007
				RE	1.2850	0.8098	0.7211	1.0000	0.6372	0.5740
	0.30	t_3	$\mathcal{N}$	Bias	0.0002	0.0002	0.0001	0.0003	0.0003	0.0002
				RE	1.2251	0.4757	0.4866	1.0000	0.5304	0.5038
	0.60	$\mathcal{N}$	t_3	Bias	0.0027	0.0014	0.0004	0.0029	0.0007	0.0015
				RE	0.9169	0.9390	0.7580	1.0000	0.7223	0.6382
					Quantile $\tau = 0.5$ — location-scale shift model					
0.0	5.00	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0012	0.0002	0.0012	0.0001	0.0023	0.0001
				RE	0.8235	2.1793	0.8235	1.0000	2.2836	1.0000
	5.00	t_3	t_3	Bias	0.0016	0.0054	0.0016	0.0005	0.0016	0.0005
				RE	1.3433	3.5334	1.3433	1.0000	2.6008	1.0000
	5.00	t_3	$\mathcal{N}$	Bias	0.0000	0.0018	0.0000	0.0007	0.0016	0.0007
				RE	0.8550	2.2129	0.8550	1.0000	2.3674	1.0000
	5.00	$\mathcal{N}$	t_3	Bias	0.0003	0.0047	0.0003	0.0001	0.0003	0.0001
				RE	1.3026	3.5993	1.3026	1.0000	2.5969	1.0000
0.4	1.15	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0001	0.0013	0.0004	0.0005	0.0006	0.0002
				RE	0.8861	1.4083	0.8495	1.0000	1.4486	0.9688
	1.25	t_3	t_3	Bias	0.0012	0.0017	0.0002	0.0004	0.0027	0.0013
				RE	1.4211	1.9722	1.2862	1.0000	1.4236	0.9181
	0.85	t_3	$\mathcal{N}$	Bias	0.0004	0.0013	0.0007	0.0013	0.0030	0.0006
				RE	1.1922	1.1522	0.9486	1.0000	1.2234	0.8803
	2.10	$\mathcal{N}$	t_3	Bias	0.0004	0.0025	0.0006	0.0001	0.0003	0.0006
				RE	1.0836	2.2118	1.1723	1.0000	1.7026	0.9714
0.8	0.35	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0007	0.0005	0.0005	0.0009	0.0002	0.0006
				RE	0.8915	0.6810	0.5851	1.0000	0.7453	0.6544
	0.45	t_3	t_3	Bias	0.0002	0.0021	0.0016	0.0001	0.0004	0.0011
				RE	1.2639	0.9453	0.8151	1.0000	0.7184	0.6304
	0.30	t_3	$\mathcal{N}$	Bias	0.0002	0.0003	0.0002	0.0006	0.0004	0.0002
				RE	1.2130	0.5704	0.5487	1.0000	0.6072	0.5611
	0.65	$\mathcal{N}$	t_3	Bias	0.0033	0.0016	0.0004	0.0021	0.0004	0.0013
				RE	0.9531	1.1496	0.9153	1.0000	0.8318	0.6951

5. Conclusions and extensions

This paper investigates a class of penalized quantile regression estimators for panel data. Assuming that the individual specific effects are drawn from distribution functions with median zero, we obtain the minimum variance estimator in the class of penalized quantile regression estimators, the analog of the Gaussian random effects in the class of penalized least squares estimators for panel data. Rather than Gaussian or Laplacian strategies, the approach seems to offer a robust alternative for λ selection because we do not need to assume spherical errors and existence of second moments.

Although our objective has been to provide a solid theoretical foundation for λ selection in quantile regression methods for panel data, several issues remain to be investigated. Under some circumstances (e.g., applications using small data sets), the researcher may not rely on estimated asymptotic variance to find the optimal degree of shrinkage. An approach based on bootstrap for λ selection is being investigated. Within the framework described above, we have been considering a device for panel data models with endogenous individual specific effects. This case arises when the independent variables and the individual specific effects are correlated. Motivated by a correlated random-effects model (Chamberlain, 1982), we have been developing a variation of the estimator to penalize uncorrelated individual effects. Monte Carlo evidence revealed that the estimator reduces the variability of the fixed-effect estimator without introducing bias, which suggests that the λ selection device proposed above may be also valid for a broader class of models.

Appendix. Proofs

The proofs refer to Knight's (1998) identities. If we denote the quantile influence function by $\psi_\tau(u) = \tau - I(u \le 0)$, for $u \neq 0$,

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau + \int_0^v (I(v \le s) - I(v \le 0))\mathrm{d}s \quad \text{(A.1)}$$

$$|u - v| - |u| = -v[I(u > 0) - I(u < 0)] + 2\int_0^v (I(u \le s) - I(u < 0))\mathrm{d}s. \quad \text{(A.2)}$$

Proof of Theorem 1. Consider the classical Gaussian random effects model (2.1). The minimization of (2.4) is equivalent to,

$$\min_{\delta_0,\boldsymbol{\delta}_1} \sum_{j=1}^J \sum_{t=1}^T \sum_{i=1}^N \omega_j \Big\{ \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j) - \delta_{0i}/\sqrt{T} - \boldsymbol{x}_{it}'\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}) - \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \Big\} + \lambda_T \left\{ \sum_{i=1}^N |\alpha_i + \delta_{0i}/\sqrt{T}| - |\alpha_i| \right\} \quad \text{(A.3)}$$

where $\xi_{it}(\tau_j) = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j)$. Following the conditions and argument of Ruppert and Carroll (1980), Koenker and Portnoy (1987), Koenker (2004), for any $(\Delta_{0i}, \Delta_1) > 0$,

$$\sup_{|\delta_{0i}|<\Delta_0, \|\delta_1\|<\Delta_1} \|v_i(\delta_{0i}, \boldsymbol{\delta}_1) - v_i(0, \mathbf{0}) - \mathbb{E}(v_i(\delta_{0i}, \boldsymbol{\delta}_1) - v_i(0, \mathbf{0}))\| = o_p(1)$$

where $\|\cdot\|$ denotes the standard Euclidean norm of a vector and,

$$v_i(\delta_{0i}, \boldsymbol{\delta}_1) = -\frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}\left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \xi_{it}(\tau_j) \right) + 2\frac{\lambda_T}{\sqrt{T}} \psi_{0.5}\left(\alpha_i + \frac{\delta_{0i}}{\sqrt{T}} \right),$$

because $\mathrm{sgn}(u) = I(u > 0) - I(u < 0) = 2\psi_{0.5}$. Taking expectation and expanding v_i under A1 and A2, we obtain

$$\begin{aligned}
&\mathbb{E}(v_i(\delta_{0i}, \boldsymbol{\delta}_1) - v_i(0, \mathbf{0})) \\
&= \mathbb{E}\Bigg(-\frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}\left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \xi_{it}(\tau_j) \right) \\
&\quad + 2\frac{\lambda_T}{\sqrt{T}} \psi_{0.5}\left(\alpha_i + \frac{\delta_{0i}}{\sqrt{T}} \right) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}\left(y_{it} - \xi_{it}(\tau_j) \right) - 2\frac{\lambda_T}{\sqrt{T}} \psi_{0.5}(\alpha_i) \Bigg) \\
&= -\frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j \left(F_{it}\left(\xi_{it}(\tau_j) + \frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) - \tau_j \right) \\
&\quad + 2\frac{\lambda_T}{\sqrt{T}} \left(G\left(-\frac{\delta_{0i}}{\sqrt{T}} \right) - \frac{1}{2} \right) \\
&= -\frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) \\
&\quad - 2\frac{\lambda_T}{\sqrt{T}} g(0) \frac{\delta_{0i}}{\sqrt{T}} + o(1)
\end{aligned}$$

where $F(\cdot)$ is the conditional distribution of y and $G(\cdot)$ is the distribution of α. While the second equality holds in part because α has a zero median distribution function, the last equality follows because G is continuous.

Clearly, $v_i(\hat{\delta}_{0i}, \hat{\delta}_1) \to 0$, and thus $\mathbb{E}(v_i(\delta_{0i}, \boldsymbol{\delta}_1) - v_i(0, \mathbf{0})) = v_i(0, \mathbf{0})$. This last expression can be written as,

$$\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) + 2\frac{\lambda_T}{\sqrt{T}} g(0) \frac{\delta_{0i}}{\sqrt{T}} \\
&= \frac{1}{\sqrt{T}} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}\left(y_{it} - \xi_{it}(\tau_j) \right) - 2\frac{\lambda_T}{\sqrt{T}} \psi_{0.5}(\alpha_i).
\end{aligned}$$

Letting $f_i = T^{-1}\sum_{j=1}^J \sum_{t=1}^T \omega_j f_{it}(\xi_{it}(\tau_j)) + 2T^{-1}\lambda_T g(0)$ and solving for δ_{0i}, we have,

$$\begin{aligned}
\frac{\delta_{0i}}{\sqrt{T}} &= f_i^{-1}\Bigg[-\frac{1}{T} \sum_{j=1}^J \sum_{t=1}^T \omega_j f_{it}(\xi_{it}(\tau_j)) \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \\
&\quad + \frac{1}{T} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}(\bullet) - 2\frac{\lambda_T}{\sqrt{T}} \frac{1}{\sqrt{T}} \psi_{0.5}(\bullet) \Bigg] + \frac{R_{Tij}}{\sqrt{T}} \\
&= -\sum_{j=1}^J \tilde{\boldsymbol{x}}_i(\tau_j)' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \frac{1}{Tf_i} \sum_{j=1}^J \sum_{t=1}^T \omega_j \psi_{\tau_j}(\bullet) \\
&\quad - 2\frac{\lambda_T}{\sqrt{T}} \frac{1}{\sqrt{T}f_i} \psi_{0.5}(\bullet) + \frac{R_{Tij}}{\sqrt{T}} \\
&= -\sum_{j=1}^J \tilde{\boldsymbol{x}}_i(\tau_j)' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + m(\psi_{\tau_j}, \psi_{0.5}, R_{Tij})
\end{aligned}$$

where $\tilde{\boldsymbol{x}}_i(\tau_j) = (Tf_i)^{-1}\sum_{t=1}^T \omega_j f_{it} \boldsymbol{x}_{it}$, and R_{Tij} is the remainder term. The first equality holds because $0 < f < \infty$ by A1 and $0 < g < \infty$ by A2. Under A5, by Lemma 1, the contribution of $m(\bullet)$ is asymptotically negligible.

The proof proceeds along similar lines as that in Koenker (2004). We follow Koenker's argument of replacing the δ_{0i}'s in (A.3) by the Bahadur representation of the individual effects in the objective function. The objective function is therefore expressed in terms of the finite dimensional slope parameter,

$$\begin{aligned}
V_{TN}(\boldsymbol{\delta}_1) &= \sum_{j=1}^J \sum_{t=1}^T \sum_{i=1}^N \omega_j \Big\{ \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j) - (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)') \\
&\quad \times \boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}) - \rho_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \Big\} \\
&\quad + \lambda_T \left\{ \sum_{i=1}^N \left| \alpha_i - \sum_{j=1}^J \tilde{\boldsymbol{x}}_i(\tau_j)' \boldsymbol{\delta}_1(\tau_j)/\sqrt{TN} \right| - |\alpha_i| \right\}.
\end{aligned}$$

Using (A.1) on the fidelity term and (A.2) on the penalty term, we decompose the objective function in four terms,

$$V_{TN}(\boldsymbol{\delta}_1) = V_{TN}^{(1)}(\boldsymbol{\delta}_1) + V_{TN}^{(2)}(\boldsymbol{\delta}_1) + V_{TN}^{(3)}(\boldsymbol{\delta}_1) + V_{TN}^{(4)}(\boldsymbol{\delta}_1)$$

where,

$$V_{TN}^{(1)}(\boldsymbol{\delta}_1) = -\sum_{j=1}^J \sum_{t=1}^T \sum_{i=1}^N \omega_j (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}) \times \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j))$$

$$V_{TN}^{(2)}(\boldsymbol{\delta}_1) = \sum_{j=1}^J \sum_{t=1}^T \sum_{i=1}^N \omega_j \int_0^{v_{itj,TN}} (I(y_{it} - \xi_{it}(\tau_j) \le s) - I(y_{it} - \xi_{it}(\tau_j) \le 0))\mathrm{d}s$$

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1) = -\lambda_T \sum_{j=1}^J \sum_{i=1}^N \tilde{\boldsymbol{x}}_i(\tau_j)'(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT})\mathrm{sgn}(\alpha_i)$$

$$V_{TN}^{(4)}(\boldsymbol{\delta}_1) = 2\lambda_T \sum_{i=1}^N \int_0^{\tilde{\boldsymbol{x}}_i'\delta_1/\sqrt{TN}} (I(\alpha_i \le s) - I(\alpha_i \le 0))\mathrm{d}s$$

with $v_{itj,TN} = (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN} = v$.

The first term is asymptotically Gaussian. Assumptions A3 and A4 implies a Lindeberg condition and we have,

$$V^{(1)}_{TN}(\boldsymbol{\delta}_1) = -\frac{1}{\sqrt{TN}}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_j(\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j) \times \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \rightsquigarrow -\boldsymbol{\delta}'_1\boldsymbol{B}.$$

We write the second term as,

$$V^{(2)}_{TN}(\boldsymbol{\delta}_1) = \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}V^{(2)}_{TN,jit}(\boldsymbol{\delta}_1) + \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(V^{(2)}_{TN,jit}(\boldsymbol{\delta}_1) - \mathbb{E}V^{(2)}_{TN,jit}(\boldsymbol{\delta}_1)\right).$$

The second term of the decomposition $V^{(2)}_{TN}(\boldsymbol{\delta}_1)$ converges in probability to a quadratic term in $\boldsymbol{\delta}_1$. Note first that,

$$\begin{aligned}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}V^{(2)}_{TN,jit}(\boldsymbol{\delta}_1) &= \frac{1}{NT}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\sqrt{TN}\int_0^{v}(F_{it}(\xi_{it}(\tau_j) + s/\sqrt{TN}) - F_{it}(\xi_{it}(\tau_j)))\mathrm{d}s\\ &= \frac{1}{2TN}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_j f_{it}(\xi_{it}(\tau_j))((\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j))^2 + o(1)\\ &\to \frac{1}{2}\boldsymbol{\delta}'_1\boldsymbol{H}_1\boldsymbol{\delta}_1.\end{aligned}$$

The second equality follows because $F(\cdot)$ is absolutely continuous and strictly increasing everywhere. Additionally, the variance of $V^{(2)}_{TN}(\boldsymbol{\delta}_1)$ converges to zero by Condition A4,

$$\mathrm{Var}(V^{(2)}_{TN}(\boldsymbol{\delta}_1)) \le \frac{1}{\sqrt{NT}}\max|(\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)| \times \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}V^{(2)}_{TN,jit}(\boldsymbol{\delta}_1) \to 0.$$

The last two terms of $V_{TN}(\boldsymbol{\delta})$ represents a decomposition of the stochastic penalty term,

$$P(\boldsymbol{\alpha}) = \lambda_T\left\{\sum_{i=1}^{N}\left|\alpha_i - \sum_{j=1}^{J}\tilde{\boldsymbol{x}}_i(\tau_j)'\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}\right| - |\alpha_i|\right\}.$$

Conditions A3 and A4 suggest a Lindeberg condition, and we have that,

$$V^{(3)}_{TN}(\boldsymbol{\delta}_1) = -\frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{N}}\sum_{j=1}^{J}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i(\tau_j)'\boldsymbol{\delta}_1(\tau_j)\mathrm{sgn}(\alpha_i) \rightsquigarrow -\lambda\boldsymbol{\delta}'_1\boldsymbol{C}.$$

The convergence in distribution is obtained by applying the Slutsky Theorem under conditions A2 and A6. Lastly, we write the fourth term as,

$$V^{(4)}_{NT}(\boldsymbol{\delta}_1) = \sum_{i=1}^{N}\mathbb{E}V^{(4)}_{NT,i}(\boldsymbol{\delta}_1) + \sum_{i=1}^{N}(V^{(4)}_{NT,i}(\boldsymbol{\delta}_1) - \mathbb{E}V^{(4)}_{NT,i}(\boldsymbol{\delta}_1)).$$

Under A2–A4 and A6, the last term $V^{(4)}_{TN}(\boldsymbol{\delta}_1)$ has a quadratic contribution to the limiting form of the objective function,

$$\begin{aligned}\mathbb{E}V^{(4)}_{TN}(\boldsymbol{\delta}_1) &= 2\frac{\lambda_T}{TN}\sum_{i=1}^{N}\int_0^{\tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1}\sqrt{NT}(G(s/\sqrt{TN}) - G(0))\mathrm{d}s\\ &= \frac{\lambda_T}{TN}\sum_{i=1}^{N}g(0)(\tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1)^2 + o(1)\\ &\to \lambda\boldsymbol{\delta}'_1\boldsymbol{H}_3\boldsymbol{\delta}_1.\end{aligned}$$

Using regularity conditions A4 and A6, we obtain,

$$\mathrm{Var}(V^{(4)}_{TN}(\boldsymbol{\delta}_1)) \le 2\frac{\lambda_T}{\sqrt{TN}}\max|\tilde{\boldsymbol{x}}_i(\tau_j)'\boldsymbol{\delta}_1(\tau_j)| \times \sum_{j=1}^{J}\sum_{i=1}^{N}\mathbb{E}V^{(4)}_{TN,ij}(\boldsymbol{\delta}_1(\tau_j)) \to 0.$$

Pollard (1991) shows that if the finite dimensional distributions of $V_{NT}(\boldsymbol{\delta}_1)$ converge weakly to those of $V_0(\boldsymbol{\delta}_1)$, and $V_0(\boldsymbol{\delta}_1)$ has a unique minimum, the minimizer of $V_{NT}(\boldsymbol{\delta}_1)$ converges in distribution to the minimizer of $V_0(\boldsymbol{\delta}_1)$. Using the argument of Pollard, because $V_{TN}(\boldsymbol{\delta}_1)$ is convex and $V_0(\boldsymbol{\delta}_1)$ has a unique minimum, we obtain that, argmin $(V_{TN}(\boldsymbol{\delta}_1)) \rightsquigarrow$ argmin $(V_0(\boldsymbol{\delta}_1))$. □

Proof of Theorem 2. The trace of the asymptotic covariance matrix is equal to,

$$\begin{aligned}&\mathrm{tr}\,\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\\ &= \sum_{j=1}^{J}\mathrm{tr}\left\{\boldsymbol{\Sigma}_1(\lambda;\tau_j)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_j,\tau_j)\boldsymbol{\Sigma}_1(\lambda;\tau_j)^{-1}\right\}\\ &= \sum_{j=1}^{J}\mathrm{tr}\,\{(\omega_j\boldsymbol{X}'\boldsymbol{M}'_j\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{X} + 2\lambda\tilde{\boldsymbol{X}}'_j\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_j)^{-1}(\Omega_{jj}\boldsymbol{X}'\boldsymbol{M}'_j\boldsymbol{M}_j\boldsymbol{X}\\ &\quad + 0.25\lambda^2\tilde{\boldsymbol{X}}'_j\tilde{\boldsymbol{X}}_j)(\omega_j\boldsymbol{X}'\boldsymbol{M}'_j\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{X} + 2\lambda\tilde{\boldsymbol{X}}'_j\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_j)^{-1}\}.\end{aligned}$$

Define the following matrices $\boldsymbol{A}^j = \omega_j(\tilde{\boldsymbol{X}}'_j\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_j)^{-1}(\boldsymbol{X}'\boldsymbol{M}'_j\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{X})$, $\boldsymbol{B}^j = \tilde{\boldsymbol{X}}'_j\boldsymbol{\Psi}\tilde{\boldsymbol{X}}_j$, $\boldsymbol{C}^j = \tilde{\boldsymbol{X}}'_j\tilde{\boldsymbol{X}}_j$, and $\boldsymbol{D}^j = \Omega_{jj}(\tilde{\boldsymbol{X}}'_j\tilde{\boldsymbol{X}}_j)^{-1}(\boldsymbol{X}'\boldsymbol{M}'_j\boldsymbol{M}_j\boldsymbol{X})$. Using Theorem 12.2.1, Graybill (1969) we have that since the matrix $\boldsymbol{C}$ is positive definite, then $\boldsymbol{C}^{-1}$ is positive definite. Using Theorem 12.2.8, Graybill (1969) we conclude that, since $\boldsymbol{C}^{-1}$ and $\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X}$ are positive definite matrices, $\boldsymbol{D}$ is also positive definite.

Replacing the matrices in the last equation gives,

$$\begin{aligned}&\mathrm{tr}\,\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\\ &= \sum_{j=1}^{J}\mathrm{tr}\left\{(\boldsymbol{A}^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{B}^j)^{-1}\boldsymbol{C}^j(\boldsymbol{D}^j + 0.25\lambda^2\boldsymbol{I})\right.\\ &\quad\left.(\boldsymbol{A}^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{B}^j)^{-1}\right\}.\end{aligned}$$

Consider the following decomposition (Rao, 1968, p. 36) for the matrices $\boldsymbol{A}^j = \boldsymbol{U}_a\boldsymbol{\Lambda}^j_a\boldsymbol{U}'_a$, where $\boldsymbol{U}$ is an orthogonal matrix, and $\boldsymbol{\Lambda}$ is a diagonal matrix that contains the characteristic roots of matrix $\boldsymbol{A}$, with a typical element ζ^{ij}_a for $i = 1,\ldots,p$. Replacing the matrices by their spectral decomposition,

$$\begin{aligned}&\mathrm{tr}\,\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\\ &= \sum_{j=1}^{J}\mathrm{tr}\,\{(\boldsymbol{U}_a\boldsymbol{\Lambda}^j_a\boldsymbol{U}'_a + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_b\boldsymbol{\Lambda}^j_b\boldsymbol{U}'_b)^{-1}\boldsymbol{U}_c\boldsymbol{\Lambda}^j_c\boldsymbol{U}'_c\\ &\quad\times(\boldsymbol{U}_d\boldsymbol{\Lambda}^j_d\boldsymbol{U}'_d + 0.25\lambda^2\boldsymbol{I})(\boldsymbol{U}_a\boldsymbol{\Lambda}^j_a\boldsymbol{U}'_a + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_b\boldsymbol{\Lambda}^j_b\boldsymbol{U}'_b)^{-1}\}.\end{aligned}$$

Note that since $\boldsymbol{U}$ is an orthogonal matrix $\boldsymbol{U}' = \boldsymbol{U}^{-1}$, and

$$\begin{aligned}(\boldsymbol{U}\boldsymbol{\Lambda}^j\boldsymbol{U}' + 2\lambda\boldsymbol{I})^{-1} &= (\boldsymbol{U}\boldsymbol{\Lambda}^j\boldsymbol{U}' + 2\lambda\boldsymbol{U}\boldsymbol{U}')^{-1} = (\boldsymbol{U}(\boldsymbol{\Lambda}^j + 2\lambda\boldsymbol{I})\boldsymbol{U}')^{-1}\\ &= \boldsymbol{U}'(\boldsymbol{\Lambda}^j + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{U},\end{aligned}$$

the last equation can be written as

$$\mathrm{tr}\,\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}$$

$$= \sum_{j=1}^{J} \operatorname{tr}\{\boldsymbol{U}_a'(\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_a\boldsymbol{U}_b'(\boldsymbol{\Lambda}_b^j)^{-1}\boldsymbol{U}_b\boldsymbol{U}_c'\boldsymbol{\Lambda}_c^j\boldsymbol{U}_c\boldsymbol{U}_d' \times (\boldsymbol{\Lambda}_d^j + 0.25\lambda^2\boldsymbol{I})\boldsymbol{U}_d\boldsymbol{U}_a'(\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_a\boldsymbol{U}_b'(\boldsymbol{\Lambda}_b^j)^{-1}\boldsymbol{U}_b\}.$$

Since the trace of $\boldsymbol{ABA}$ is equal to the trace of $\boldsymbol{AAB}$,

$$\operatorname{tr}\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1} = \sum_{j=1}^{J} \operatorname{tr}\{\boldsymbol{U}_a'\boldsymbol{U}_a(\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_b'\boldsymbol{U}_b(\boldsymbol{\Lambda}_b^j)^{-1}\boldsymbol{U}_c'\boldsymbol{U}_c\boldsymbol{\Lambda}_c^j\boldsymbol{U}_d'\boldsymbol{U}_d \times (\boldsymbol{\Lambda}_d^j + 0.25\lambda^2\boldsymbol{I})\boldsymbol{U}_a'\boldsymbol{U}_a(\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_b'\boldsymbol{U}_b(\boldsymbol{\Lambda}_b^j)^{-1}\}.$$

Consequently, since $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$, the equation is now

$$\operatorname{tr}\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1} = \sum_{j=1}^{J} \operatorname{tr}\left\{(\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_b^j)^{-1}\boldsymbol{\Lambda}_c^j(\boldsymbol{\Lambda}_d^j + 0.25\lambda^2\boldsymbol{I}) \times (\boldsymbol{\Lambda}_a^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_b^j)^{-1}\right\}$$
$$= \sum_{j=1}^{J}\sum_{i=1}^{p} \frac{\zeta_c^{ij}(\zeta_d^{ij} + 0.25\lambda^2)}{(\zeta_b^{ij}(\zeta_a^{ij} + 2\lambda))^2} = \sum_{j=1}^{J}\sum_{i=1}^{p} \pi(\lambda)^{ij} = \Pi(\lambda).$$

We now have a simple optimization problem as a function of λ. The discontinuities of the objective function are ruled out since the matrix $\boldsymbol{B}^j$ is positive definite, which implies that the eigenvalues $\zeta_b^{ij} > 0$ for all i, j.

Since $\boldsymbol{A}^j$, $\boldsymbol{C}^j$ and $\boldsymbol{D}^j$ are positive definite, their eigenvalues ζ_a^{ij}, ζ_c^{ij} and ζ_d^{ij} are positive for all i, j. Using Lemma 4, $\pi(\lambda)^{ij}$ is convex in λ in $\mathcal{A}^{ij}$ for all i, j. Since the sets $\mathcal{A}^{ij}$ are a decreasing sequence of sets, the functions $\pi(\lambda)^{ij}$ are also convex, using Lemma 5, in $\lambda \in \mathcal{D} = \cap_{ij}\mathcal{A}^{ij}$. Since the sum of convex functions is also convex, $\Pi(\lambda)$ is also convex in $\lambda \in \mathcal{D}$.

Therefore, the function $\Pi(\lambda) : \mathcal{D} \to \mathbb{R}_+$ is a continuous strictly convex function defined on a non-empty, compact set (Lemma 3). These sufficient conditions implies that $\Pi(\lambda)$ has a unique minimizer, $\exists\lambda^* \in \mathcal{D}$ such that $\Pi(\lambda^*) < \Pi(\lambda)$ for all $\lambda \in \mathcal{D}$. □

Proof of Corollary 1. The convexity of $V_{NT}(\boldsymbol{\delta}_1)$ implies that $\hat{\boldsymbol{\delta}}_1$ converges in distribution to the minimizer of $V_0(\boldsymbol{\delta}_1)$. By Theorem 1, the limiting form of the objective function is equal to,

$$V_0(\boldsymbol{\delta}_1) = -\boldsymbol{\delta}_1'(\boldsymbol{B} + \lambda\boldsymbol{C}) + \frac{1}{2}\boldsymbol{\delta}_1'(\boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3)\boldsymbol{\delta}_1.$$

Thus its minimizer is $\boldsymbol{\delta}_1 = (\boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3)^{-1}(\boldsymbol{B} + \lambda\boldsymbol{C})$. By A2, $\mathbb{E}\boldsymbol{\delta}_1 = \boldsymbol{0}$, and

$$\operatorname{Var}(\boldsymbol{\delta}_1) = (\boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3)^{-1}(\boldsymbol{H}_0 + \lambda^2\boldsymbol{H}_2)(\boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3)^{-1} = \boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}.$$

The result follows since $\boldsymbol{\Gamma}_1(\lambda)$ is positive definite for all $\lambda \in \mathbb{R}_+$, which is sufficient condition for a minimum. □

Proof of Corollary 2. It is immediate from previous results from considering one τ. □

Proof of Corollary 3. It is immediate by fixing λ to λ^*. □

Proof of Corollary 4. The jk-th diagonal element of the asymptotic covariance matrix is

$$\left(\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\right)_{jk,jk} = \left(\boldsymbol{\Sigma}_1(\lambda;\tau_j)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_j,\tau_j)\boldsymbol{\Sigma}_1(\lambda;\tau_j)^{-1}\right)_{k,k}$$
$$= (\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k + 2\lambda\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk})^{-1}(\Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k + 0.25\lambda^2\tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk})(\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k + 2\lambda\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk})^{-1}$$
$$= \frac{(\Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k + 0.25\lambda^2\tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk})}{(\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k + 2\lambda\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk})^2}.$$

Let $A_{jk} = \Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k$, $B_{jk} = \tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk}$, $C_{jk} = \omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k$, and $D_{jk} = \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk}$. We write the variance of $\hat{\beta}_k(\tau_j)$ as

$$\left(\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\right)_{jk,jk} = \frac{A_{jk} + 0.25\lambda^2 B_{jk}}{(C_{jk} + 2\lambda D_{jk})^2} = \Pi(\lambda).$$

Note that the discontinuities are ruled out since $C_{jk} > 0$, and $D_{jk} > 0$. The first derivative of the function $\Pi(\lambda)$ with respect to λ is

$$\frac{\partial\Pi(\lambda)}{\partial\lambda} = \frac{0.5\lambda B_{jk}}{(C_{jk} + 2\lambda D_{jk})^2} - \frac{4D_{jk}(A_{jk} + 0.25\lambda^2 B_{jk})}{(C_{jk} + 2\lambda D_{jk})^3} = \frac{0.5C_{jk}B_{jk}\lambda - 4A_{jk}D_{jk}}{(C_{jk} + 2\lambda D_{jk})^3} = 0.$$

Therefore, the minimum variance quantile quantile regression estimator for the unobserved effects model is defined for

$$\lambda^* = \frac{8A_{jk}D_{jk}}{C_{jk}B_{jk}} = \frac{8\Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk}}{\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk}}. \tag{A.4}$$

We need to show that the second order condition evaluated at λ^* is positive.

$$\frac{\partial^2\Pi(\lambda)}{\partial\lambda^2} = \frac{24A_{jk}(D_{jk})^2 + B_{jk}C_{jk}(0.5C_{jk} - 2\lambda D_{jk})}{(C_{jk} + 2\lambda D_{jk})^4} = \frac{(B_{jk})^4(C_{jk})^4(0.5B_{jk}(C_{jk})^2 + 8A_{jk}(D_{jk})^2)}{(B_{jk}(C_{jk})^2 + 16A_{jk}(D_{jk})^2)^4} > 0.$$

The objective function is strictly convex in λ since $A_{jk} > 0$ and $B_{jk} > 0$, which is a sufficient condition for the uniqueness of λ^*. □

Lemma 1. *Under the conditions of Theorem 1, the terms of the Bahadur representation of the individual effect that do not depend on the main parameter of interest are $o_p(1)$.*

Proof. We need to show that the arguments of

$$m(\psi_{\tau_j}, \psi_{0.5}, R_{Tij}) = \sum_{j=1}^{J} \underbrace{\frac{1}{Tf_i}\sum_{t=1}^{T}\omega_j\psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j))}_{\bar{\psi}_i(\tau_j)} - 2\frac{\lambda_T}{\sqrt{T}}\underbrace{\frac{1}{\sqrt{T}f_i}\psi_{0.5}(\alpha_i)}_{\bar{\psi}_i(0.5)} + \frac{R_{Tij}}{\sqrt{T}}$$

are asymptotically negligible. First, we need to show that $\sum_j \bar{\psi}_i(\tau_j) \to 0$ as $T \to \infty$. Note that for a fixed number of quantiles, by regularity condition A6,

$$f_i = \frac{1}{T}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}(\xi_{it}(\tau_j)) + \frac{\lambda_T}{\sqrt{T}}\frac{2g}{\sqrt{T}} \to \mathbb{E}f_i + \lambda\, 0 = \mathbb{E}f_i.$$

Using the familiar Chebyshev inequality,

$$P\left\{\left|\sum_{j=1}^{J}\bar{\psi}_i(\tau_j) - \mathbb{E}\sum_{j=1}^{J}\bar{\psi}_i(\tau_j)\right| \geq \epsilon\right\} \leq \frac{1}{\epsilon^2}\mathbb{E}\left(\sum_{j=1}^{J}\bar{\psi}_i(\tau_j) - \mathbb{E}\sum_{j=1}^{J}\bar{\psi}_i(\tau_j)\right)^2$$

$$= \frac{1}{\epsilon^2}\mathbb{E}\left((\bar{\psi}_i(\tau_1) - \mathbb{E}\bar{\psi}_i(\tau_1)) + \cdots + (\bar{\psi}_i(\tau_J) - \mathbb{E}\bar{\psi}_i(\tau_J))\right)^2$$
$$= \frac{1}{\epsilon^2}\mathbb{E}\left(\bar{\psi}_i(\tau_1) + \cdots + \bar{\psi}_i(\tau_J)\right)^2.$$

The last equality holds because trivially $\mathbb{E}\bar{\psi}_i(\tau_j) = 0$ for all $j \in [1, J]$. Also, by conditional independence assumed in A1, we have,

$$\mathrm{Var}(\bar{\psi}_i(\tau_j)) = \frac{1}{(Tf_i)^2}\mathrm{Var}\left(\sum_{t=1}^{T}\omega_j\psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j))\right)$$
$$= \frac{\omega_j^2}{(Tf_i)^2}\sum_{t=1}^{T}\mathrm{Var}(\psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)))$$
$$= \frac{\omega_j^2}{(Tf_i)^2}\sum_{t=1}^{T}\tau_j(1-\tau_j) = \frac{\omega_j^2\tau_j(1-\tau_j)}{Tf_i^2}.$$

Moreover, by independence,

$$\mathbb{E}(\bar{\psi}_i(\tau_j)\bar{\psi}_i(\tau_k)) = \frac{1}{(Tf_i)^2}\mathbb{E}\left(\left(\sum_{t=1}^{T}\omega_j\psi_{\tau_j}(\cdot)\right)\left(\sum_{t=1}^{T}\omega_k\psi_{\tau_k}(\cdot)\right)\right)$$
$$= \frac{\omega_j\omega_k}{(Tf_i)^2}\sum_{t=1}^{T}\mathbb{E}\left(\psi_{\tau_j}(\cdot)\psi_{\tau_k}(\cdot)\right) = \frac{\omega_j\omega_k\tau_j(1-\tau_k)}{Tf_i^2}.$$

Therefore,

$$P\left\{\left|\sum_{j=1}^{J}\bar{\psi}_{it}(\tau_j) - \mathbb{E}\sum_{j=1}^{J}\bar{\psi}_{it}(\tau_j)\right| \geq \epsilon\right\}$$
$$\leq \frac{\sum_{j=1}^{J}\omega_j^2\tau_j(1-\tau_j) + \sum_{j\neq h}^{J}\omega_j\omega_h\tau_j(1-\tau_h)}{Tf_i^2\epsilon^2},$$

which tends to zero as T tends to infinity. The first term is therefore asymptotically negligible.

Second, we need to show that $\bar{\psi}_i(0.5) \to 0$. Again using Chebyshev inequality,

$$P\left\{\left|\bar{\psi}_i(0.5) - \mathbb{E}\bar{\psi}_i(0.5)\right| \geq \epsilon\right\} \leq \frac{1}{\epsilon^2}\mathbb{E}(\bar{\psi}_i(0.5) - \mathbb{E}\bar{\psi}_i(0.5))^2$$
$$= \frac{1}{\epsilon^2}\mathbb{E}(\bar{\psi}_i(0.5))^2.$$

Similarly, as before, $\mathrm{Var}(\bar{\psi}_i(0.5)) = 0.5 \times 0.5/Tf_i^2$. Consequently, the second term is also asymptotically negligible, because $\lambda_T/\sqrt{T} \to \lambda$ by A6, and,

$$P\left\{\left|\bar{\psi}_i(0.5) - \mathbb{E}\bar{\psi}_i(0.5)\right| \geq \epsilon\right\} \leq \frac{1}{4T\epsilon^2 f_i^2} \to 0.$$

Lastly, the remainder term is asymptotically negligible by A5. It can be shown that the condition is needed to use Corollary 3.2.3 in van der Vaart and Wellner (1996) for uniform convergence of a simple argmax process (see, e.g., Wei and He (2006) for a similar approach). As demonstrated in Koenker (2004), the contribution of the remainder term is $o_p(1)$. This completes the proof. □

Lemma 2. *Let $\mathcal{A} = [0, \bar{\lambda}] \subset \mathbb{R}_+$ where $\bar{\lambda} = 12\zeta_d/\zeta_a + 0.25\zeta_a - \epsilon > 0$ for positive constants ζ_a, ζ_d, and ϵ. Then, the set $\mathcal{A}$ is non-empty, closed and bounded.*

Lemma 3. *Let $\mathcal{A}_{(n)}$ be a decreasing sequence of sets (e.g., $\mathcal{A}_{(1)} \supseteq \mathcal{A}_{(2)} \supseteq \cdots \supseteq \mathcal{A}_{(N)}$). Then, the set $\mathcal{D} = \cap_{i=1}^{N}\mathcal{A}_{(n)}$ is non-empty, closed and bounded.*

Proof. Note that $\mathcal{D} = \cap_{i=1}^{N}\mathcal{A}_{(n)} = \mathcal{A}_{(N)}$. By Lemma 2, $\mathcal{D}$ is non-empty, closed and bounded. □

Lemma 4. *Let* (a) *$\zeta_a, \zeta_b, \zeta_c, \zeta_d$ be positive constants, and* (b) *the parameter $\lambda \in \mathcal{A}$. Then, the rational function $\pi(\lambda) : \mathcal{A} \to \mathbb{R}_+$,*

$$\pi(\lambda) = \frac{\zeta_c(\zeta_d + 0.25\lambda^2)}{(\zeta_b(\zeta_a + 2\lambda))^2}$$

is a $\mathcal{C}^\infty$ differentiable function, strictly convex in λ.

Proof. The discontinuities are ruled out since $\zeta_b > 0$. The function π is a rational function, and every rational function is continuous. The function is convex if $\partial^2\pi(\lambda)/\partial\lambda^2$ is strictly positive. The first derivative of the function $\pi(\lambda)$ with respect to λ is

$$\frac{\partial\pi(\lambda)}{\partial\lambda} = \frac{0.5\zeta_c\lambda}{\zeta_b^2(\zeta_a + 2\lambda)^2} - \frac{4\zeta_c(\zeta_d + 0.25\lambda^2)}{\zeta_b^2(\zeta_a + 2\lambda)^3} = \frac{\zeta_c(0.5\zeta_a\lambda - 4\zeta_d)}{\zeta_b^2(\zeta_a + 2\lambda)^3}.$$

The second derivative is equal to,

$$\frac{\partial^2\pi(\lambda)}{\partial\lambda^2} = \frac{0.5\zeta_c}{\zeta_b^2(\zeta_a + 2\lambda)^2} - \frac{4\zeta_c\lambda}{\zeta_b^2(\zeta_a + 2\lambda)^3} + \frac{24\zeta_c(\zeta_d + 0.25\lambda^2)}{\zeta_b^2(\zeta_a + 2\lambda)^4}$$
$$= \frac{\zeta_c(0.5\zeta_a^2 + 24\zeta_d - 2\zeta_a\lambda)}{\zeta_b^2(\zeta_a + 2\lambda)^4} > 0.$$

Since $0 < \lambda < 12\zeta_d/\zeta_a + 0.25\zeta_a$, the function is strictly convex over the domain of π. □

Lemma 5. *If $\pi(\lambda)$ is a strictly convex function over $\mathcal{A} \subset \mathbb{R}_+$, the function $\pi(\lambda)$ is also strictly convex over $\mathcal{D} \subset \mathcal{A}$.*

Proof. Suppose not; then $\pi(\alpha\lambda_1 + (1-\alpha)\lambda_2) \geq \alpha\pi(\lambda_1) + (1-\alpha)\pi(\lambda_2)$ for all $\lambda_1, \lambda_2 \in \mathcal{D}$, and $0 \leq \alpha \leq 1$. The points λ_1, λ_2 are also in $\mathcal{A}$, but this is a contradiction since $\pi(\lambda)$ is strictly convex over $\mathcal{A}$. □

References

Abrevaya, J., Dahl, C., 2008. The effects of smoking and prenatal care on birth outcomes: evidence from quantile regression estimation on panel data. Journal of Business and Economics Statistics 26 (4), 379–397.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory.

Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in colombia: evidence from a randomized natural experience. American Economic Review 92 (5), 1535–1558.

Baltagi, B., 1981. Pooling: an experimental study of alternative testing and estimation procedures in a two-way error component model. Journal of Econometrics 17, 21–49.

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18, 5–46.

Donoho, D., Chen, S., Saunders, M., 1998. Atomic decomposition by basis pursuit. SIAM Journal of Scientific Computing 20, 33–61.

Geraci, M., Bottai, M., 2007. Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics 8 (1), 140–154.

Graybill, F., 1969. Introduction to Matrices with Applications in Statistics. Wadsworth.

Hoerl, A., Kennard, R., 1988. Ridge regression. Encyclopedia of Statistical Sciences 8, 129–136.

Horowitz, J., Markatou, M., 1996. Semiparametric estimation of regression models for panel data. Review of Economics Studies 63, 145–168.

Knight, K., 1998. Limiting distributions for L_1 regression estimators under general conditions. Annals of Statistics 26, 755–770.

Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.

Koenker, R., 2005. Quantile Regression. Cambridge University Press.

Koenker, R., 2006. Quantreg, R package version 3.90. www.r-project.org.

Koenker, R., Bassett, G., 1982. Tests of linear hypothesis and ℓ_1 estimation. Econometrica 50, 1577–1584.

Koenker, R., Hallock, K., 2000. Quantile regression: an introduction, mimeo, Department of Economics, University of Illinois at Urbana-Champaign.

Koenker, R., Portnoy, S., 1987. L-estimation for linear models. Journal of the American Statistical Association 82, 851–857.

Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70, 1583–1612.

Kooperberg, C., Stone, C., 1991. A study of logspline density estimation. Computational Statistics and Data Analysis 12, 327–347.

Nerlove, M., 1971. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. Econometrica 39 (2), 359–382.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.
Rao, C., 1968. Linear Statistical Inference and Its Applications. Wiley, New York.
Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects (with discussion). Statistical Science 6 (1), 15–51.
Ruppert, D., Carroll, R., 1980. Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 75, 828–838.
Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press.
Schwartz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice and Visualization. Wiley and Sons, New York.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B 58, 267–288.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. Springer, New York.
Wei, Y., He, X., 2006. Conditional growth charts. Annals of Statistics 34 (5), 2069–2097.
Zou, H., Hastie, T., Tibshirani, R., 2007. On the Degrees of Freedom of the Lasso. Annals of Statistics 35 (5), 2173–2192.