PENALIZED QUANTILE REGRESSION ESTIMATION FOR A MODEL WITH ENDOGENOUS INDIVIDUAL EFFECTS*

CARLOS LAMARCHE†

Abstract. This paper proposes a penalized quantile regression estimator for panel data that explicitly considers individual heterogeneity associated with the covariates. We provide conditions under which the estimator is asymptotically unbiased and Gaussian, thus the harshness of the penalization can be determined by minimizing estimated variance. We investigate finite sample and asymptotic performance in terms of quadratic loss in a class of quantile regression estimators. The evidence suggests that the penalized approach can significantly reduce the variability of existing quantile regression estimators for panel data models with endogenous individual effects, without introducing bias. Three empirical applications of the method illustrate the approach.

Keywords: Endogeneity; Quantile regression; Penalty method; Data dependent shrinkage.

1. Introduction

The estimation of longitudinal models in economics and statistics often requires the use of a robust technique that allows the possibility of estimating covariate effects at different quantiles of the conditional distribution of the response variable, while controlling for individual heterogeneity. This paper is concerned with the estimation of quantile regression functions for panel data models that explicitly considers individual time-invariant heterogeneity associated with the independent variables. Specifically, we will consider the following model,

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \alpha + u \tag{1.1}$$

$$\tau \neq P(\alpha \leq 0|\boldsymbol{x}) \tag{1.2}$$

where y is the response variable, $\boldsymbol{x}$ is a vector of independent variables, u is the error term, and τ is the median quantile. Equation (1.1) represents the classical linear random effects model. Equation (1.2) suggests that the individual specific effect α may be drawn

*This Version: June 13, 2008. First Version: March 11, 2008. The author would like to thank Roger Koenker, Steve Portnoy and Kevin Grier for helpful comments, as well as seminar participants at the University of Oklahoma and the University of Illinois at Urbana-Champaign. The College of Art and Sciences and the Supercomputing Center for Education and Research at the University of Oklahoma provided financial and computing support. The R software for this paper and results are available upon request.

†Department of Economics, University of Oklahoma. 321 Hester Hall, 729 Elm Avenue, Norman, OK 73019. Tel.: +1 405 325 5857. Email: lamarche@ou.edu.

from a non-zero median distribution function, and that the location of this distribution is not independent of $\boldsymbol{x}$. In this framework, it is natural to consider estimating directly a vector of individual effects, but the procedure inflates the variability of the estimates of the covariate effects and does not allow estimation of time invariant effects. In both Angrist et al. (2002, 2006) longitudinal analysis of a voucher program and De Silva et al. (2008) study of a state policy affecting procurement auctions of construction contracts, it is not possible to estimate time-invariant treatment indicators. This paper presents a quantile regression approach that both deals with endogeneity and allows identification of time-invariant effects.

Research on quantile regression models for panel data is relatively new. Koenker (2004) introduces a class of penalized quantile regression estimators proposing to estimate directly a vector of individual effects. The estimation of these parameters increases the variability of the estimates of the $\boldsymbol{\beta}$'s, but regularization, or shrinkage reduces the inflation effect. Lamarche (2006) provides conditions under which it is possible to obtain the minimum variance estimator in the class of penalized estimators, the analog of the GLS in the class of penalized least squares estimators for panel data. Geraci and Bottai (2006) and Abrevaya and Dahl (2008) propose different approaches. The first paper uses a likelihood approach under asymmetric Laplace distributions, and the second paper considers the "correlated random effects" model of Chamberlain (1982, 1984). While shrinkage produces bias in the estimate of the covariate effect when $P(\alpha \leq 0|\boldsymbol{x})$ is not independent of $\boldsymbol{x}$, Chamberlain's framework adapted for quantile regression inflates the variability of the estimate of the covariate effect even for a small number of observations on each subject.

This paper proposes a penalized quantile regression method for estimating (1.1)-(1.2), considering that the location of the distribution of α can be represented as a function of the covariates $h(\boldsymbol{x})$. The approach improves the performance of existing quantile regression methods for panel data by both allowing correlation between $\boldsymbol{x}$ and α and increasing the precision of the estimate of the $\boldsymbol{\beta}$'s. We provide conditions under which the estimator is asymptotically unbiased and Gaussian, thus the harshness of the penalization can be determined by minimizing estimated variance. Monte Carlo evidence reveals that the estimator can eliminate the bias that arises when the endogeneity of the observables is ignored, and significantly reduce the variability of existing methods that give unbiased results.

Our approach is closely connected to the correlated random effects framework and related models that have been largely considered in empirical economics (e.g, Jakubson (1988), Ashenfelter and Krueger (1994), Carey (1997), Ashenfelter and Rouse (1998), Krashinsky (2004), Ziliak (2003), among others). We illustrate the use of the method in three applications. The first example uses a subsample of genetically identical twins from Ashenfelter and Krueger (1994) to estimate the return of education. In the second application, we investigate the distributional effect of background risk on wealth considering the framework developed in Carroll and Samwick (1998) and Ziliak (2003). Lastly, we estimate the intertemporal substitution elasticity of labor-supply using the British Household Panel Survey (BHPS). These examples show interesting differences among quantile regression approaches for panel data, and demonstrate that the approach offers the possibility of estimating quantile models with suspected endogenous variables while achieving better performance relative to existing methods.

The next section presents the model and estimator. Section 3 studies the asymptotic properties of the estimator and Section 4 offers Monte-Carlo evidence. Section 5 demonstrates how the penalized estimator can be obtained and used in empirical applications. Section 6 provides conclusions.

2. Model and Estimator

Consider the classical Gaussian random effects model

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

where y_{it} is the dependent variable, $\boldsymbol{x}_{it} = (1, x_{it,2}, ..., x_{it,p})'$ is the vector of independent variables, the α_i's are unobservable time-invariant effects, and u_{it} is the error term. We allow the variables α_i and $\boldsymbol{x}_{it}$ to be stochastically dependent by considering the individual effect to be drawn from a conditional distribution function with location $h(\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\gamma}$, with $\boldsymbol{x}_i = (\boldsymbol{x}_{it})_t$. This may be seen within the classical context of Chamberlain (1982, 1984) framework leading to a more familiar representation of the endogenous individual effects,

$$\alpha_i = \boldsymbol{x}_i'\boldsymbol{\gamma} + a_i.$$

The individual effect a_i is distributed as G_i, and by definition, this effect is uncorrelated with the independent variables.

Although standard panel data methods offer the possibility of estimating conditional mean models while controlling for individual heterogeneity, until recently, few papers have estimated conditional quantile functions with individual specific effects. This paper considers individual heterogeneity associated with the covariates estimating conditional quantile functions of the form,

$$Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{x}_i, a_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \boldsymbol{x}_i'\boldsymbol{\gamma}(\tau) + a_i,$$

for all quantiles τ's in the interval (0,1). The parameter of interest is $\boldsymbol{\beta}(\tau)$. The individual specific effect a_i is a location shift effect on the conditional quantiles of the response as in Koenker (2004). We will also estimate an alternative version of this model assuming that the individual effect α_i does not represent a distributional shift, since it is unrealistic to estimate it when the number of observations on each individual is small. In the empirical sections, we will occasionally impose the condition that the covariate effect represents both a location shift $\boldsymbol{\gamma}$ and a distributional shift $\boldsymbol{\gamma}(\tau)$.

2.1. **Estimation.** Naturally, the estimation of the individual effects a_i's in the quantile regression model increases the variability of the estimators of the covariate effects. As explained below, we use regularization, or shrinkage of these individual effects to deal with this problem.

There is an enormous amount of work in statistics and lately in econometrics dealing with regularization in a wide spectrum of problems including estimation of models with a large number of parameters (see, e.g., Tibshirani (1996), Koenker (2004), Horowitz and Lee (2007), Carrasco, Florens and Renault (2007), Chen (2007); see also Bickel and Li (2006) for a survey in statistics). For estimation of the quantile model with individual effects, we consider the estimator that solves,

$$\min \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j \rho_{\tau_j}(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau_j) - \boldsymbol{x}'_i\boldsymbol{\gamma}(\tau_j) - a_i) + \lambda Pen(\boldsymbol{a}),$$

where $\rho_{\tau_j}(u) = u(\tau_j - I(u \leq 0))$ is the quantile loss function, ω_j is a relative weight given to the j-th quantile, and λ is the *Tikhonov* regularization parameter or tuning parameter. The function $Pen(\boldsymbol{a})$ is a ℓ_1 penalty term that could be defined as $\|\boldsymbol{a} - \boldsymbol{a}^*\|_1$, where $\boldsymbol{a}^*$ may be close to the unknown location of the distribution. In the Chamberlain's model by definition a_i has zero mean, so we made use of this information defining the penalty term as

$$Pen(\boldsymbol{a}) = \|\boldsymbol{a}\|_1.$$

The estimation of the individual effects increases the variability of the estimators of the covariate effects, but this penalty term that shrinks the fixed effects estimator of the a_i's toward zero helps to reduce the inflation effect without sacrificing bias. The estimation method solves a version of the penalized estimator introduced by Koenker (2004), considering the following design matrix,

$$\begin{bmatrix} \text{diag}(\omega) \otimes \boldsymbol{X} & \text{diag}(\omega) \otimes \boldsymbol{Z}\boldsymbol{D} & \omega \otimes \boldsymbol{Z} \\ \boldsymbol{0} & \boldsymbol{0} & \lambda \boldsymbol{I} \end{bmatrix},$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & \boldsymbol{x}'_{11} \\ 1 & \boldsymbol{x}'_{12} \\ \vdots & \\ 1 & \boldsymbol{x}'_{NT} \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} \boldsymbol{x}'_{11} & \boldsymbol{x}'_{12} & \cdots & \boldsymbol{x}'_{1T} \\ \boldsymbol{x}'_{21} & \boldsymbol{x}'_{22} & \cdots & \boldsymbol{x}'_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}'_{N1} & \boldsymbol{x}'_{N2} & \cdots & \boldsymbol{x}'_{NT} \end{bmatrix}; \boldsymbol{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

For identification when $\lambda \to 0$, we will need the standard conditions on restricting the estimation to $n-1$ individual specific effects.

As in any regularization problem, the selection of the tuning parameter λ is of fundamental interest. In non parametrics, the tuning parameter λ is typically selected by generalized cross validation, in ridge regression by minimizing minimum squared error, and in classical panel data models by maximum likelihood or generalized least squares (Ruppert, Wand, and Carroll 2003). In general the ℓ_1 penalty function $\|\boldsymbol{a}\|_1$ does not achieve unbiasedness, but in the case of exchangeable a_i's with zero-median distribution function, shrinkage improves the slope's performance without sacrificing bias. The heuristics of finding the optimal value of the tuning parameter suggests, in the present framework, to find,

$$\hat{\lambda} = \arg\inf \{tr\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\},$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is the covariance matrix of the slope parameter. The matrix may be estimated using the standard bootstrap and alternative resampling methods for quantile regression that has been investigated, among others, by Buchinsky (1995), Hahn (1995), Horowitz (1998). The empirical covariance matrix $\boldsymbol{\Sigma}$ can be easily computed given λ and B bootstrap estimates $\{\hat{\boldsymbol{\beta}}^*(\boldsymbol{\tau}), \hat{\boldsymbol{\gamma}}^*(\boldsymbol{\tau}), \hat{\boldsymbol{a}}^*\}$. These bootstrap estimates are obtained using block or panel bootstrap, that is, sampling pairs $\{(\boldsymbol{y}_i, \boldsymbol{x}_i) : i = 1, ..., N\}$ with replacement. Next section provides a more rigorous approach for λ selection based on minimizing asymptotic variance.

3. Asymptotic Properties

The asymptotic theory of the penalized estimator can be developed using the existing asymptotic results on panel data (e.g., Koenker 2004). We will employ the following regularity conditions:

A 1. The variables y_{it} are independent with conditional distribution $F_{Y_{it}}$, and continuous densities f_{it} uniformly bounded away from 0 and ∞ at the points $\xi_{it}(\tau_j)$ for $j = 1, \ldots, J$, $t = 1, \ldots, T$ and $i = 1, \ldots, N$.

A 2. The random variables a_i are exchangeable, identically, and independently distributed with unconditional distribution function G_i with median zero, and continuous densities g_i for $i = 1, \ldots, N$.

A 3. There exist positive definite matrices $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$, and $\boldsymbol{\Sigma}_3$ such that

$$\boldsymbol{\Sigma}_0 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{TN} \begin{bmatrix} \Omega_{11}\boldsymbol{X}'\boldsymbol{W}_1'\boldsymbol{W}_1\boldsymbol{X} & \ldots & \Omega_{1J}\boldsymbol{X}'\boldsymbol{W}_1'\boldsymbol{W}_J\boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \Omega_{1J}\boldsymbol{X}'\boldsymbol{W}_J'\boldsymbol{W}_1\boldsymbol{X} & \ldots & \Omega_{JJ}\boldsymbol{X}'\boldsymbol{W}_J'\boldsymbol{W}_J\boldsymbol{X} \end{bmatrix}$$

$$\mathbf{\Sigma}_1 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{TN} \begin{bmatrix} \omega_1 \boldsymbol{X}'\boldsymbol{W}_1'\boldsymbol{\Upsilon}_1\boldsymbol{W}_1\boldsymbol{X} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \omega_J \boldsymbol{X}'\boldsymbol{W}_J'\boldsymbol{\Upsilon}_J\boldsymbol{W}_J\boldsymbol{X} \end{bmatrix}$$

$$\mathbf{\Sigma}_2 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{\Omega_m}{TN} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{P}_1'\boldsymbol{P}_1\boldsymbol{X} & \dots & \boldsymbol{X}'\boldsymbol{P}_1'\boldsymbol{P}_J\boldsymbol{X} \\ \vdots & \ddots & \vdots \\ \boldsymbol{X}'\boldsymbol{P}_J'\boldsymbol{P}_1\boldsymbol{X} & \dots & \boldsymbol{X}'\boldsymbol{P}_J'\boldsymbol{P}_J\boldsymbol{X} \end{bmatrix}$$

$$\mathbf{\Sigma}_3 = \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{NT} \begin{bmatrix} \boldsymbol{X}'\boldsymbol{P}_1'\mathbf{\Psi}\boldsymbol{P}_1\boldsymbol{X} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boldsymbol{X}'\boldsymbol{P}_J'\mathbf{\Psi}\boldsymbol{P}_J\boldsymbol{X} \end{bmatrix}$$

where $\Omega_{kl} = \omega_k(\tau_k \wedge \tau_l - \tau_k\tau_l)\omega_l$ and $\Omega_m = \tau_m(1-\tau_m)$ for the median τ_m; $\boldsymbol{W}_j = \boldsymbol{I} - \boldsymbol{Z}\boldsymbol{P}_j$, $\boldsymbol{P}_j = (\boldsymbol{Z}'\boldsymbol{\Upsilon}_j\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Upsilon}_j$, $\boldsymbol{\Upsilon}_j = \mathbf{\Phi}_j(\boldsymbol{I} - \mathbf{\Lambda}_j)$, $\mathbf{\Phi}_j = \mathrm{diag}(f_{it}(\xi_{it}(\tau_j)))$, $\mathbf{\Lambda}_j = \mathrm{diag}(\boldsymbol{x}_i'(\boldsymbol{D}'\boldsymbol{Z}'\mathbf{\Phi}_j\boldsymbol{Z}\boldsymbol{D})^{-1}\mathbf{\Phi}_{ij}\boldsymbol{x}_i)$, and $\mathbf{\Psi} = \mathrm{diag}(g_i(0))$.

A 4. $\max||\boldsymbol{x}_{it}||/\sqrt{TN} \to 0$ and $\max||\boldsymbol{x}_i||/\sqrt{T} \to 0$.

Condition A1 ensures a well defined asymptotic behavior of the estimator by imposing variability around the J conditional quantiles. Condition A2 may be seen as an extension of Chamberlain's (1982) orthogonality assumption between a_i and $\boldsymbol{x}_{it}$. This condition always holds in the case of correlated random effects. The existence of the limiting form of the positive definite matrices, assumed in A3, is needed to invoke the Lindeberg-Feller Central Limit Theorem. Condition A4 is important for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function.

3.1. **Asymtotic Normality.** Under the previous conditions, it is possible to show that the penalized estimator is asymptotically unbiased and Gaussian. The argument given in the Appendix follows closely Koenker (2004) overcoming the difficulties associated with the infinite dimension of $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ by concentrating out these effects in the objective function. The idea is to focus on the limiting behavior of the penalized slope $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$.

Theorem 1. *Under regularity conditions A1-A4, and provided that $N^c/T \to 0$ for some $c > 0$ and $\lambda_T/\sqrt{T} \to \lambda \geq 0$, the penalized quantile regression estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ is asymptotically normally distributed with mean $\boldsymbol{\beta}(\boldsymbol{\tau})$ and covariance matrix,*

$$(\mathbf{\Sigma}_1 + \lambda\mathbf{\Sigma}_3)^{-1}(\mathbf{\Sigma}_0 + \lambda^2\mathbf{\Sigma}_2)(\mathbf{\Sigma}_1 + \lambda\mathbf{\Sigma}_3)^{-1}.$$

Remark 1. When the a's are exchangeable and drawn from a conditional distribution function with location zero, shrinkage that forces some individual specific effect estimates $\hat{a}$'s to be zero does not impose bias and affects the performance of the estimator of the parameter of interest $\boldsymbol{\beta}(\boldsymbol{\tau})$.

Remark 2. The estimation of the asymptotic covariance matrix can be accomplished by obtaining estimates of the conditional density f at the conditional quantile $\xi(\tau)$ and the density of the individual effects $g(0)$. The estimation of $f(\xi(\tau))$ in iid and non-iid settings requires the use of standard quantile regression methods, considering the conditional quantile function evaluated at λ equal to zero, $\xi(\tau, 0)$. The interested reader will find in Koenker (2005) detailed explanations on the existing approaches. The estimation of $g(0)$ can be performed considering a sample of normalized individual effects estimates $\{\hat{a}_1(0), \hat{a}_2(0), ..., \hat{a}_N(0)\}$ and classical kernel methods, $(1/(Nh_N))\sum_{i=1}^{N} K(\hat{a}_i(0)/h_N)$, where h_N is a bandwidth and $\hat{a}_i(0)$ is the "unpunished" estimate of the individual effect a_i.

3.2. **Asymptotic Relative Efficiency.** We compare the performance of three estimators of $\boldsymbol{\beta}$: the estimator that penalizes uncorrelated individual effects $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$, the estimator that penalizes correlated individual effects $\tilde{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$, and the quantile regression estimator for the correlated random effects model $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, 0)$. The estimator $\tilde{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda)$ is similar to the estimator considered in Koenker (2004) when the location of the distribution of the *iid* α_i's is different than zero, and $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, 0)$ is similar to the estimator considered in Abrevaya and Dahl (2008) replacing the time effects by individual effects. Below, we facilitate the comparison restricting attention to estimators in the class of penalized quantile regression estimators, considering the median quantile τ and normalized matrices $\boldsymbol{A}$ denoted by $\boldsymbol{A}_o$.

The estimator that shrinks the individual specific effects α_i converges in distribution to the normal random variable $\boldsymbol{v}$,

$$\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau, \lambda) - \boldsymbol{\beta}(\tau)) \rightsquigarrow (\boldsymbol{H}_1 + \lambda \boldsymbol{H}_3)^{-1}(\boldsymbol{B} + \lambda \boldsymbol{C}) = \boldsymbol{v},$$

where $\boldsymbol{B}$ is a Gaussian vector independent of $\boldsymbol{C}$ with covariance $\boldsymbol{H}_0 = \tau(1-\tau)\boldsymbol{X}_o'\boldsymbol{M}_o'\boldsymbol{M}_o\boldsymbol{X}_o$, $\boldsymbol{C}$ is a Gaussian vector with covariance $\boldsymbol{H}_2 = \tau(1-\tau)\boldsymbol{X}_o'\boldsymbol{P}_{\boldsymbol{\Phi}}'\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}_o$, $\boldsymbol{H}_1 = \boldsymbol{X}_o'\boldsymbol{M}_o'\boldsymbol{\Phi}_o\boldsymbol{M}_o\boldsymbol{X}_o$, and $\boldsymbol{H}_3 = \boldsymbol{X}_o'\boldsymbol{P}_{\boldsymbol{\Phi}}'\boldsymbol{\Psi}_o\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}_o$. The weighted projection matrices are $\boldsymbol{M}_o = \boldsymbol{I} - \boldsymbol{Z}_o\boldsymbol{P}_{\boldsymbol{\Phi}}$ and $\boldsymbol{P}_{\boldsymbol{\Phi}} = (\boldsymbol{Z}_o'\boldsymbol{\Phi}_o\boldsymbol{Z}_o)^{-1}\boldsymbol{Z}_o'\boldsymbol{\Phi}_o$. While $\boldsymbol{B}$ is a mean zero variable because the error term and the covariates are stochastically independent, the vector $\boldsymbol{C}$ is an non-zero mean vector because the sign of $\boldsymbol{\alpha}$ and the independent variables are correlated. Consequently, we have that $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$ is asymptotically biased for positive values of the tuning parameter λ:

$$\text{Abias}(\tilde{\boldsymbol{\beta}}(\tau, \lambda)) = \mathbb{E}\boldsymbol{v} = (\boldsymbol{H}_1 + \lambda \boldsymbol{H}_3)^{-1}\mathbb{E}(\boldsymbol{B} + \lambda \boldsymbol{C}) = (\boldsymbol{H}_1 + \lambda \boldsymbol{H}_3)^{-1}\lambda \boldsymbol{S}_o,$$

with $\mathbb{E}\boldsymbol{C} = \boldsymbol{S}_o$. Note that the bias may be not negligible even for small values of λ and the absolute value asymptotic bias increases at a decreasing rate as $\lambda \to \infty$. Moreover, the asymptotic variance of the penalized estimator is convex in λ and equal to,

$$\text{Avar}(\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) = (\boldsymbol{H}_1 + \lambda \boldsymbol{H}_3)^{-1}(\boldsymbol{H}_0 + \lambda^2 \boldsymbol{H}_2)(\boldsymbol{H}_1 + \lambda \boldsymbol{H}_3)^{-1}.$$

We now turn to the estimator that shrinks the "pure" random effects a_i's. Theorem 1 establishes that the estimator is asymptotically unbiased and has asymptotic variance equal to,

$$\text{Avar}(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) = (\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0 + \lambda^2 \boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1},$$

where $\boldsymbol{\Sigma}_0 = \tau(1-\tau)\boldsymbol{X}_o'\boldsymbol{W}_o'\boldsymbol{W}_o\boldsymbol{X}_o$, $\boldsymbol{\Sigma}_1 = \boldsymbol{X}_o'\boldsymbol{W}_o'\boldsymbol{\Upsilon}_o\boldsymbol{W}_o\boldsymbol{X}_o$, $\boldsymbol{\Sigma}_2 = \tau(1-\tau)\boldsymbol{X}_o'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}$, $\boldsymbol{\Sigma}_3 = \boldsymbol{X}_o'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{\Psi}_o\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}_o$. The weighted projection matrices are $\boldsymbol{W}_o = \boldsymbol{I}_o - \boldsymbol{Z}_o\boldsymbol{P}_{\boldsymbol{\Upsilon}}$ and $\boldsymbol{P}_{\boldsymbol{\Upsilon}} = (\boldsymbol{Z}_o'\boldsymbol{\Upsilon}_o\boldsymbol{Z}_o)^{-1}\boldsymbol{Z}_o'\boldsymbol{\Upsilon}_o$. In what follows, we let $\boldsymbol{L} = \boldsymbol{M}_o\boldsymbol{X}_o$ and $\boldsymbol{R} = \boldsymbol{W}_o\boldsymbol{X}_o$.

The matrices $\boldsymbol{P}_{\boldsymbol{\Phi}}$ and $\boldsymbol{P}_{\boldsymbol{\Upsilon}}$ produce the same transformation. Note that,

$$\begin{aligned}
\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X} &= (\boldsymbol{Z}'\boldsymbol{\Upsilon}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Upsilon}\boldsymbol{X} \\
&= (\boldsymbol{Z}'\boldsymbol{\Phi}(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{X} \\
&= (T\boldsymbol{Z}'\boldsymbol{\Phi}(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z})^{-1}T\boldsymbol{Z}'\boldsymbol{\Phi}(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{X} \\
&= (\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{X} \\
&= (\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}(\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z})^{-1}\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{\Lambda})\boldsymbol{Z}\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{X} \\
&= (\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{X} = \boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}.
\end{aligned}$$

Since the previous result implies that $\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}_o$ is equal to $\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}_o$, we have that $\boldsymbol{H}_0 = \boldsymbol{\Sigma}_0 = \boldsymbol{J}_0$, $\boldsymbol{H}_2 = \boldsymbol{\Sigma}_2 = \boldsymbol{J}_2$, and $\boldsymbol{H}_3 = \boldsymbol{\Sigma}_3 = \boldsymbol{J}_3$. Notice that the conditional density of α_i at the median is equal to the unconditional density of a_i at zero. Therefore, the asymptotic relative efficiency between $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ and $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ is determined by

$$\begin{aligned}
\boldsymbol{H}_1 - \boldsymbol{\Sigma}_1 &= \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{L} - \boldsymbol{R}'\boldsymbol{\Upsilon}\boldsymbol{R} = \boldsymbol{L}'(\boldsymbol{\Phi}-\boldsymbol{\Upsilon})\boldsymbol{L} \\
&= \boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{L} = ||\boldsymbol{L}||^2_{\boldsymbol{\Phi}\boldsymbol{\Lambda}} > \boldsymbol{0}.
\end{aligned}$$

with the inequality indicating that $\boldsymbol{H}_1 - \boldsymbol{\Sigma}_1$ is positive definite and implying that the asymptotic variance of the penalized estimator $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ is smaller than the asymptotic variance of $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ for all λ's in $\mathbb{R}_+$.

Theorem 2. *Under the conditions of Theorem 1, for $\lambda \in (0,\infty)$, the penalized estimator that shrinks endogenous individual effects, $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$, and the penalized estimator that shrinks exogenous individual effects, $\hat{\boldsymbol{\beta}}(\tau,\lambda)$, have covariance matrices,*

$$\begin{aligned}
Avar(\sqrt{NT}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) &= (\boldsymbol{H}_1 + \lambda \boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2 \boldsymbol{J}_2)(\boldsymbol{H}_1 + \lambda \boldsymbol{J}_3)^{-1}, \\
Avar(\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) &= (\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{J}_3)^{-1}(\boldsymbol{J}_0 + \lambda^2 \boldsymbol{J}_2)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{J}_3)^{-1},
\end{aligned}$$

and $Avar(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) < Avar(\hat{\boldsymbol{\beta}}(\tau,\lambda))$. Also,

$$|Abias(\tilde{\boldsymbol{\beta}}(\tau,\lambda))| > |Abias(\hat{\boldsymbol{\beta}}(\tau,\lambda))| = |Abias(\hat{\boldsymbol{\beta}}(\tau,0))| = \boldsymbol{0}.$$

Remark 3. The result could be interpreted in terms of asymptotic mean squared error (AMSE). Note that although $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ is asymptotically biased, it may have asymptotically significantly smaller variance than the unbiased estimators $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ and $\hat{\boldsymbol{\beta}}(\tau,0)$. We find that the trace of the asymptotic mean squared error is:

$$\begin{aligned}
\mathrm{AMSE}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) &= \sum_{i=1}^{p} \frac{\zeta_{\tilde{c}}^{i}(\zeta_{\tilde{d}}^{i}+\lambda^2)}{(\zeta_{\tilde{a}}^{i}(\zeta_{\tilde{b}}^{i}+\lambda))^2} + \frac{\bar{\zeta}_{S_o}\lambda^2}{(\bar{\zeta}_{\tilde{a}}(\bar{\zeta}_{\tilde{b}}+\lambda))^2} \\
\mathrm{AMSE}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) &= \sum_{i=1}^{p} \frac{\zeta_{\hat{c}}^{i}(\zeta_{\hat{d}}^{i}+\lambda^2)}{(\zeta_{\hat{a}}^{i}(\zeta_{\hat{b}}^{i}+\lambda))^2} \\
\mathrm{AMSE}(\hat{\boldsymbol{\beta}}(\tau,0)) &= \sum_{i=1}^{p} \frac{\zeta_{\hat{c}}^{i}\zeta_{\hat{d}}^{i}}{(\zeta_{\hat{a}}^{i}\zeta_{\hat{b}}^{i})^2}
\end{aligned}$$

where the ζ's are positive eigenvalues of the matrices $\tilde{\boldsymbol{A}} = \hat{\boldsymbol{A}} = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{\Psi}\boldsymbol{P}\boldsymbol{X}$, $\tilde{\boldsymbol{C}} = \hat{\boldsymbol{C}} = \tau(1-\tau)\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}$, $\tilde{\boldsymbol{D}} = \hat{\boldsymbol{D}} = \tau(1-\tau)\tilde{\boldsymbol{C}}^{-1}\boldsymbol{L}'\boldsymbol{L}$, $\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{A}}^{-1}\boldsymbol{L}'\boldsymbol{\Phi}\boldsymbol{L}$ and $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{L}'\boldsymbol{\Upsilon}\boldsymbol{L}$. Note that while $\zeta_{\tilde{a}}^{i} = \zeta_{\hat{a}}^{i}$, $\zeta_{\tilde{c}}^{i} = \zeta_{\hat{c}}^{i}$ and $\zeta_{\tilde{d}}^{i} = \zeta_{\hat{d}}^{i}$. Moreover, by Weyl's monotonicity principle of eigenvalues (Bhatia 1997),

$$\zeta_{\tilde{b}}^{i} > \zeta_{\hat{b}}^{i} \quad \forall i.$$

The AMSE of the penalized estimator $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ is a function of the eigenvalues of the asymptotic quadratic bias term. The largest eigenvalue is defined as $\bar{\zeta}_{S_o} = \max\{\zeta_{S_o}^{1},\ldots,\zeta_{S_o}^{p}\}$ and, $\bar{\zeta}_{\tilde{a}}$ and $\bar{\zeta}_{\tilde{b}}$ are the corresponding eigenvalues. From the expressions, it is immediately apparent than for $\bar{\zeta}_{S_o}$ sufficiently small,

$$\mathrm{AMSE}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) \le \mathrm{AMSE}(\hat{\boldsymbol{\beta}}(\tau,\lambda)),$$

but the inequality is reversed if the bias and the tuning parameter are large. For λ sufficiently small, we have that

$$\mathrm{AMSE}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) \le \mathrm{AMSE}(\hat{\boldsymbol{\beta}}(\tau,0)),$$

suggesting that shrinking the individual effects a's is worthwhile. It is obvious that a critical aspect of this estimation technique is the ability to do λ selection, so we investigate this issue in the next section.

3.3. **Optimal Shrinkage.** We consider choosing λ to minimize AMSE, which for the case of $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau},\lambda)$ implies choosing λ to minimize asymptotic variance. The primary objective is now to show that the trace of the asymptotic covariance matrix is convex in λ, therefore a unique value of λ exists.

Theorem 3. *Under the conditions of Theorem 1, there exists a unique variance minimizing parameter,*

$$\lambda^* = \arg\min\{tr(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1)(\boldsymbol{\Sigma}_1+\lambda\boldsymbol{\Sigma}_3)^{-1}(\boldsymbol{\Sigma}_0+\lambda^2\boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1+\lambda\boldsymbol{\Sigma}_3)^{-1}\}.$$

Remark 4. Theorem 3 demonstrates that it is possible to obtain an optimal tuning parameter defined as the minimizer of the trace of the asymptotic covariance matrix. Note that the selection of λ^* is not sensitive to scale effects because we consider normalized asymptotic variances $\text{Avar}(\hat{\beta}_k(\tau,\lambda))/\text{Avar}(\hat{\beta}_k(\tau,0))$.

Remark 5. Wand (1999) derives a closed-form asymptotic approximation for the optimal degree of smoothing in penalized spline regression. In the case of $p = 1$, it may be advantageous to consider the optimal tuning parameter as the minimizer of the j-th diagonal element,

$$\frac{\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}\boldsymbol{e} + \lambda^2\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}\boldsymbol{e}}{(\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{\Upsilon}\boldsymbol{W}\boldsymbol{X}\boldsymbol{e} + \lambda\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}\boldsymbol{e})^2},$$

where $\boldsymbol{e}$ is a vector containing an indicator variable for the covariate. For values of $\lambda \to 0$, a small increase in the shrinkage parameter reduces the variance of the estimator, and for values $\lambda \to \infty$, the variance increases. The variance is continuous and strictly convex, therefore λ^* exists and is,

$$\lambda^* = \frac{\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{\Upsilon}\boldsymbol{W}\boldsymbol{X}\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}'\boldsymbol{P}_{\boldsymbol{\Upsilon}}\boldsymbol{X}\boldsymbol{e}}.$$

In the case of a_i and u_{it} distributed as iid Gaussian variables, the densities g_i and f_{it} are constant and equal to g and f. Note that $\boldsymbol{W}\boldsymbol{X} = \boldsymbol{M}\boldsymbol{X}$, with $\boldsymbol{M}$ being idempotent. Under these conditions, the optimal tuning parameter is $\lambda^* = cg/f$ for a constant $c > 1$.

Remark 6. Under the previous regularity conditions and considering the methods described in Remark 2, a "plug-in" estimator $\hat{\lambda}$ consistently estimates the optimal degree of shrinkage λ^*, therefore the feasible estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau},\hat{\lambda})$ should be indistinguishable from the unfeasible estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau},\lambda^*)$ as the sample size increases.

4. Monte Carlo

This section reports the results of several simulation experiments designed to evaluate the performance of the method in finite samples. First, we will briefly investigate the bias and variance of the penalized estimator in models with endogenous individual effects. Second, we will contrast the performance of the penalized quantile regression estimator for the correlated random effects model with classical least squares estimators and quantile regression estimators. Lastly, we will evaluate the efficiency of the penalized estimator relative to existing approaches for panel data quantile regression.

4.1. **Experiment Design.** We generate the dependent variable considering the following equations,

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + \alpha_i + (1 + \delta x_{it}) u_{it} \\ x_{it} &= \pi \mu_i + v_{it} \\ \alpha_i &= \gamma_0 + \gamma_1 x_{i1} + ... + \gamma_T x_{iT} + a_i \end{aligned}$$

where u_{it} and a_i are iid Gaussian variables. While the independent variable x_{it} is also generated as iid Gaussian variables in the location shift model, it is distributed as χ^2_3 in the location-shift model to avoid quantile curves that cross each other. The results for the location shift model were similar to the results for the location-scale shift model, so we will only report estimates for $\delta = 0$. In all the variant of the models reported on the tables below, the β's are assumed to be zero, the γ's are $0.5/T$ representing the Mundlak-Chamberlain case, and π is set to be 2.5.

4.2. **Results.** We start reporting results on the performance of the penalized quantile regression estimators when the correlation between the independent variables and the individual specific effect is small $\gamma = \{0.02, 0.05, 0.10\}$ and $\beta_1 = 1$. We consider a data set with $N = 100$ and $T = 5$. The panels of Figure 1 report the bias and variance percentage change of the penalized estimator, which penalizes the α_i's (PQR) and the penalized estimator that penalizes the a_i's (PCQR). The panels on the left show that the PQR estimator is biased, and its bias starts to increase, as we increase the harshness of the penalization. For $\gamma = 0.02$ for instance, the bias of the slope PQR estimator achieves 5 percent for λ approximately equal to 1, while the bias of the PCQR is zero. The right panels reveal that the variance of the estimators decrease first and then increase, but there are significant differences in variance reduction. By carefully choosing λ to be 1, the variance of the slope PQR estimator can be reduced more than 25 percent, while the variance of the PCQR is reduced by 2 percent. The evidence also shows that the variance compression does not dramatically depend on the correlation between α_i and x_{it}, but the bias of the PQR does. We observe a proportional change in the bias of the PQR estimator, as it changes from 5 to 25 percent when γ increases from 0.02 to 0.10.

We expand the design of the experiment considering several sample sizes $N = \{100, 250, 500\}$ and $T = \{2, 4, 12\}$, and the random variables a_i and u_{it} to be distributed as Gaussian and t-student with 3 degrees of freedom. Considering these models in Table 4.1, we compare the performance of the following estimators: (1) the ordinary least squares (OLS); (2) the generalized least squares (GLS); (3) the pooled quantile regression estimator (QR); (4) Koenker's (2004) penalized quantile regression estimator (PQR); (5) the penalized quantile regression estimator for the correlated random effects model (PCQR); (6) Abrevaya and Dahl's (2008) quantile regression estimator (CQR).

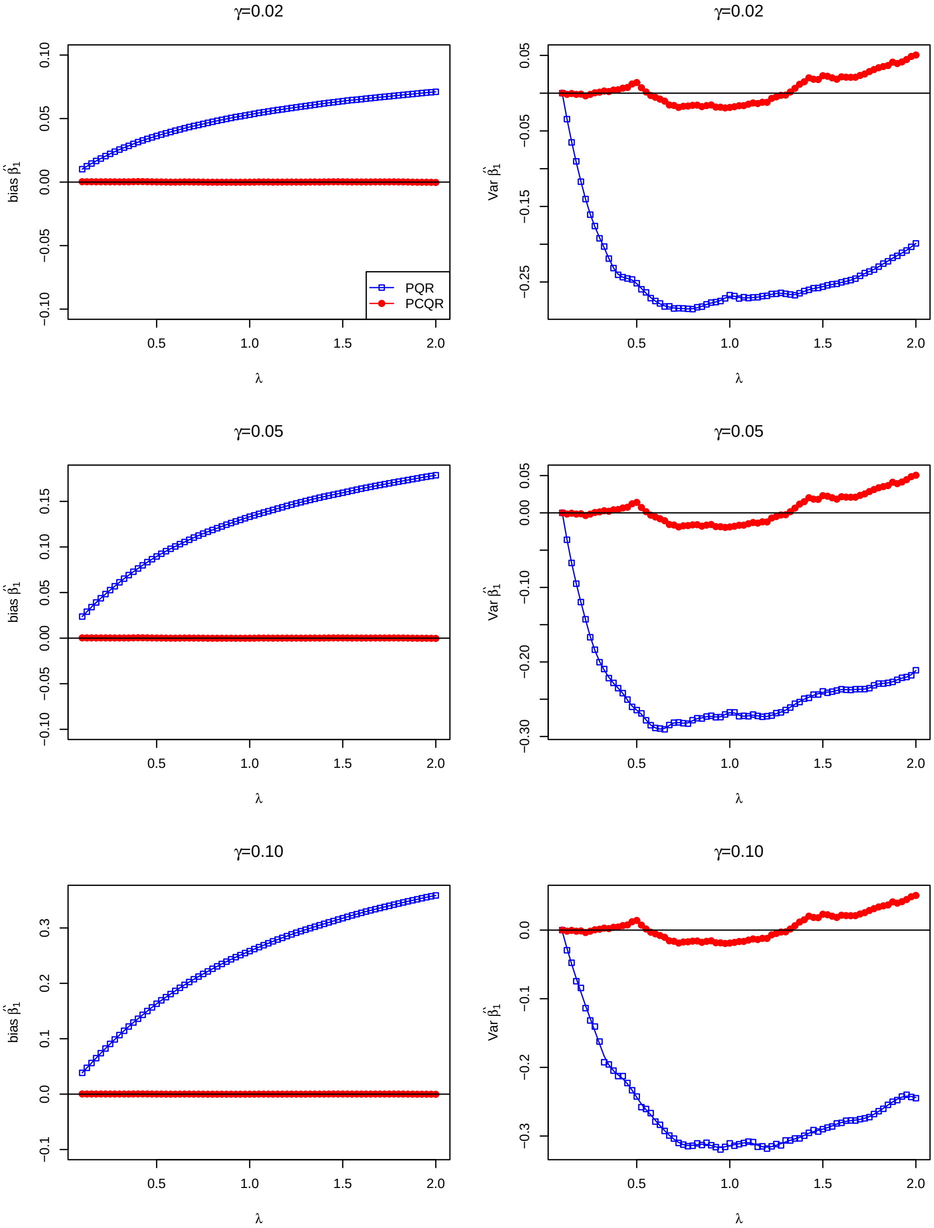

FIGURE 4.1. *Quantile regression estimation for the correlated effects model when γ's are positive and small. The left panels show the bias and the right panels the variance percentage change. PQR stands for the estimator that penalizes "endogenous" individual effects, and PCQR stands for the estimator that penalizes "exogenous" individual effects. Each dot represents a statistic based on 400 randomly generated samples.*

			Estimators						
			Least Squares		Quantile Regression				
			OLS	GLS	QR	PQR	PCQRd	PCQRl	CQR
N	T	Statistics	$\mathcal{N}(0,1)$ distributions						
100	2	Bias	0.4578	0.3900	0.4573	0.4070	-0.0086	-0.0104	-0.0077
		Std Dev	0.0435	0.0408	0.0523	0.0494	0.1144	0.1063	0.1395
100	4	Bias	0.4355	0.2877	0.4360	0.2986	0.0038	0.0046	0.0001
		Std Dev	0.0395	0.0345	0.0470	0.0458	0.0719	0.0638	0.0812
100	12	Bias	0.4235	0.1496	0.4229	0.1397	0.0019	0.0021	0.0023
		Std Dev	0.0391	0.0263	0.0415	0.0314	0.0387	0.0346	0.0421
250	2	Bias	0.4594	0.3955	0.4599	0.4109	-0.0001	-0.0011	-0.0014
		Std Dev	0.0285	0.0272	0.0337	0.0326	0.0757	0.0706	0.0887
250	4	Bias	0.4407	0.3006	0.4412	0.3112	0.0006	-0.0003	0.0007
		Std Dev	0.0243	0.0212	0.0283	0.0287	0.0416	0.0374	0.0515
250	12	Bias	0.4322	0.1650	0.4324	0.1513	0.0004	-0.0003	0.0007
		Std Dev	0.0238	0.0151	0.0252	0.0186	0.0246	0.0219	0.0279
500	2	Bias	0.4644	0.4077	0.4646	0.4213	0.0024	0.0012	0.0050
		Std Dev	0.0190	0.0181	0.0233	0.0229	0.0529	0.0468	0.0623
500	4	Bias	0.4490	0.3175	0.4495	0.3289	-0.0014	-0.0018	-0.0010
		Std Dev	0.0175	0.0154	0.0197	0.0193	0.0310	0.0275	0.0389
500	12	Bias	0.4404	0.1801	0.4407	0.1654	-0.0001	-0.0003	-0.0008
		Std Dev	0.0151	0.0106	0.0165	0.0139	0.0163	0.0145	0.0178
N	T	Statistics	t_3 distributions						
100	2	Bias	0.4608	0.3933	0.4600	0.4110	0.0012	0.0013	0.0056
		Std Dev	0.0818	0.0770	0.0716	0.0673	0.1393	0.1340	0.1739
100	4	Bias	0.4312	0.2819	0.4323	0.2975	-0.0045	-0.0048	-0.0038
		Std Dev	0.0705	0.0604	0.0569	0.0532	0.0847	0.0750	0.0971
100	12	Bias	0.4242	0.1483	0.4212	0.1403	-0.0015	-0.0010	-0.0006
		Std Dev	0.0703	0.0409	0.0518	0.0364	0.0421	0.0390	0.0530
250	2	Bias	0.4601	0.3969	0.4602	0.4123	-0.0016	-0.0006	-0.0036
		Std Dev	0.0480	0.0441	0.0426	0.0419	0.0866	0.0830	0.1003
250	4	Bias	0.4423	0.3008	0.4409	0.3150	-0.0047	-0.0032	-0.0050
		Std Dev	0.0448	0.0382	0.0353	0.0309	0.0507	0.0484	0.0630
250	12	Bias	0.4313	0.1651	0.4316	0.1570	0.0011	0.0009	0.0014
		Std Dev	0.0404	0.0274	0.0294	0.0237	0.0250	0.0241	0.0318
500	2	Bias	0.4644	0.4081	0.4644	0.4226	0.0014	0.0031	0.0015
		Std Dev	0.0338	0.0322	0.0294	0.0284	0.0633	0.0576	0.0728
500	4	Bias	0.4498	0.3179	0.4498	0.3338	0.0007	-0.0005	0.0007
		Std Dev	0.0300	0.0270	0.0238	0.0238	0.0385	0.0349	0.0455
500	12	Bias	0.4385	0.1792	0.4389	0.1696	-0.0004	-0.0002	-0.0005
		Std Dev	0.0264	0.0187	0.0217	0.0168	0.0203	0.0183	0.0240

TABLE 4.1. *Bias and Standard Deviation of the Estimators. Both the penalizes estimator quantile regression estimator (PQR) and the penalized estimator for the correlated random effects (PCQR) are defined for λ^*. The table presents two versions of the estimator: PCQRd estimates the model considering $\gamma(\tau)$ and PCQRl considers $\gamma(\tau) = \gamma$.*

			Asymptotic Theory			Bootstrap		
			Quantiles					
			0.25	0.5	0.75	0.25	0.5	0.75
N	T	Statistics	$\mathcal{N}(0,1)$ distributions					
100	2	Bias	0.0104	0.0120	0.0079	0.0081	0.0192	0.0038
		Std Err	0.1467	0.1142	0.1397	0.1414	0.1258	0.1356
		RE	0.9329	0.7922	0.8887	0.8991	0.8722	0.8629
100	4	Bias	0.0008	-0.0002	-0.0009	0.0015	0.0003	-0.0009
		Std Err	0.0780	0.0680	0.0786	0.0773	0.0674	0.0772
		RE	0.9264	0.8577	0.9046	0.9179	0.8499	0.8893
100	12	Bias	0.0017	-0.0006	-0.0039	0.0021	-0.0007	-0.0030
		Std Err	0.0416	0.0377	0.0408	0.0409	0.0384	0.0414
		RE	0.8437	0.8644	0.9251	0.8292	0.8794	0.9408
250	2	Bias	-0.0009	-0.0051	-0.0030	0.0015	-0.0065	-0.0043
		Std Err	0.0965	0.0768	0.0925	0.0986	0.0789	0.0925
		RE	1.0213	0.8724	0.9620	1.0433	0.8961	0.9618
250	4	Bias	0.0026	0.0001	-0.0023	0.0035	0.0006	-0.0021
		Std Err	0.0493	0.0435	0.0529	0.0504	0.0466	0.0493
		RE	0.8926	0.8440	0.9620	0.9129	0.9044	0.8971
250	12	Bias	0.0006	-0.0006	0.0003	0.0005	-0.0007	0.0003
		Std Err	0.0248	0.0220	0.0259	0.0256	0.0224	0.0260
		RE	0.7700	0.8781	0.9360	0.7960	0.8965	0.9378
500	2	Bias	-0.0020	-0.0021	-0.0006	-0.0033	0.0003	-0.0007
		Std Err	0.0656	0.0520	0.0687	0.0629	0.0544	0.0646
		RE	0.9662	0.8378	1.0303	0.9268	0.8770	0.9682
500	4	Bias	-0.0009	-0.0010	-0.0020	-0.0009	-0.0007	-0.0013
		Std Err	0.0362	0.0322	0.0378	0.0367	0.0330	0.0360
		RE	0.8538	0.8946	0.9336	0.8664	0.9179	0.8889
500	12	Bias	-0.0004	-0.0005	-0.0009	-0.0005	-0.0006	-0.0010
		Std Err	0.0193	0.0169	0.0182	0.0195	0.0167	0.0180
		RE	0.8862	0.8646	0.8913	0.8980	0.8572	0.8783
N	T	Statistics	t_3 distributions					
100	2	Bias	-0.0002	-0.0013	0.0003	-0.0011	0.0044	0.0045
		Std Err	0.1872	0.1429	0.1796	0.1795	0.1499	0.1781
		RE	0.8604	0.8185	0.8196	0.8253	0.8585	0.8125
100	4	Bias	0.0032	-0.0011	0.0027	-0.0004	0.0011	0.0039
		Std Err	0.1036	0.0829	0.0967	0.1041	0.0820	0.0965
		RE	0.8685	0.7743	0.8300	0.8724	0.7662	0.8287
100	12	Bias	-0.0020	-0.0007	-0.0025	-0.0015	-0.0008	-0.0031
		Std Err	0.0497	0.0407	0.0513	0.0500	0.0412	0.0515
		RE	0.8613	0.8188	0.8052	0.8672	0.8289	0.8081
250	2	Bias	-0.0005	-0.0032	-0.0034	0.0008	-0.0018	-0.0075
		Std Err	0.1170	0.0885	0.1085	0.1229	0.0873	0.1149
		RE	0.9126	0.8753	0.8900	0.9591	0.8640	0.9421
250	4	Bias	0.0032	0.0012	0.0013	0.0047	0.0009	0.0002
		Std Err	0.0630	0.0512	0.0625	0.0639	0.0515	0.0625
		RE	0.7982	0.8908	0.7962	0.8096	0.8974	0.7970
250	12	Bias	-0.0005	0.0006	0.0020	0.0002	0.0006	0.0020
		Std Err	0.0319	0.0254	0.0305	0.0323	0.0253	0.0309
		RE	0.7947	0.8038	0.8302	0.8056	0.8020	0.8411
500	2	Bias	-0.0004	0.0031	0.0067	0.0008	0.0025	0.0055
		Std Err	0.0865	0.0655	0.0815	0.0834	0.0666	0.0796
		RE	0.9838	0.8774	0.9130	0.9493	0.8912	0.8919
500	4	Bias	-0.0003	-0.0011	0.0007	-0.0003	-0.0026	0.0001
		Std Err	0.0442	0.0398	0.0465	0.0472	0.0385	0.0476
		RE	0.8412	0.8854	0.8971	0.8975	0.8563	0.9175
500	12	Bias	0.0015	-0.0004	0.0005	0.0013	-0.0004	0.0003
		Std Err	0.0229	0.0192	0.0238	0.0236	0.0195	0.0241
		RE	0.7753	0.8504	0.8590	0.8017	0.8642	0.8701

TABLE 4.2. *Feasible PCQRd estimation. The table considers λ^* estimated using asymptotic and bootstrapped variance. RE stands for the relative efficiency of Abrevaya and Dahl's estimator (PCQ) relative to the penalized estimator (PCQRd).*

			Asymptotic Theory			Bootstrap		
					Quantiles			
			0.25	0.5	0.75	0.25	0.5	0.75
N	T	Statistics			$\mathcal{N}(0,1)$ distributions			
100	2	Bias	0.0072	0.0103	0.0098	0.0122	0.0152	0.0123
		Std Err	0.1099	0.1075	0.1083	0.1099	0.1094	0.1080
		RE	0.6987	0.7456	0.6889	0.6989	0.7587	0.6871
100	4	Bias	0.0004	-0.0008	0.0001	0.0006	-0.0003	0.0003
		Std Err	0.0611	0.0601	0.0607	0.0609	0.0601	0.0609
		RE	0.7257	0.7585	0.6995	0.7239	0.7581	0.7011
100	12	Bias	-0.0002	-0.0006	-0.0003	-0.0003	-0.0004	-0.0005
		Std Err	0.0347	0.0348	0.0349	0.0351	0.0349	0.0351
		RE	0.7042	0.7980	0.7914	0.7115	0.8002	0.7976
250	2	Bias	-0.0033	-0.0036	-0.0054	-0.0024	-0.0039	-0.0049
		Std Err	0.0727	0.0706	0.0720	0.0680	0.0712	0.0698
		RE	0.7694	0.8019	0.7484	0.7192	0.8089	0.7253
250	4	Bias	-0.0003	-0.0002	-0.0013	0.0011	0.0010	-0.0002
		Std Err	0.0403	0.0388	0.0404	0.0413	0.0400	0.0409
		RE	0.7287	0.7539	0.7347	0.7472	0.7767	0.7435
250	12	Bias	0.0003	-0.0004	-0.0001	0.0003	-0.0004	-0.0003
		Std Err	0.0201	0.0201	0.0203	0.0203	0.0202	0.0202
		RE	0.6254	0.8038	0.7324	0.6316	0.8090	0.7294
500	2	Bias	-0.0015	-0.0020	-0.0022	-0.0011	-0.0017	-0.0019
		Std Err	0.0496	0.0475	0.0488	0.0492	0.0479	0.0487
		RE	0.7304	0.7646	0.7312	0.7249	0.7718	0.7302
500	4	Bias	-0.0011	-0.0012	-0.0004	-0.0009	-0.0009	0.0000
		Std Err	0.0293	0.0292	0.0295	0.0292	0.0295	0.0294
		RE	0.6910	0.8105	0.7296	0.6893	0.8202	0.7258
500	12	Bias	-0.0005	-0.0007	-0.0005	-0.0007	-0.0008	-0.0007
		Std Err	0.0154	0.0152	0.0154	0.0152	0.0150	0.0152
		RE	0.7076	0.7784	0.7529	0.7016	0.7694	0.7453
N	T	Statistics			t_3 distributions			
100	2	Bias	0.0002	0.0010	0.0007	0.0027	0.0034	0.0015
		Std Err	0.1445	0.1391	0.1421	0.1477	0.1399	0.1443
		RE	0.6642	0.7970	0.6483	0.6790	0.8016	0.6584
100	4	Bias	0.0017	0.0013	0.0024	0.0015	0.0018	0.0013
		Std Err	0.0796	0.0765	0.0779	0.0793	0.0772	0.0775
		RE	0.6675	0.7142	0.6691	0.6643	0.7214	0.6654
100	12	Bias	-0.0024	-0.0009	-0.0021	-0.0020	-0.0010	-0.0020
		Std Err	0.0400	0.0381	0.0386	0.0408	0.0382	0.0388
		RE	0.6936	0.7665	0.6062	0.7072	0.7701	0.6094
250	2	Bias	-0.0010	-0.0030	-0.0037	-0.0016	-0.0032	-0.0026
		Std Err	0.0846	0.0820	0.0817	0.0836	0.0821	0.0830
		RE	0.6602	0.8110	0.6700	0.6523	0.8117	0.6807
250	4	Bias	0.0030	0.0029	0.0012	0.0028	0.0015	0.0006
		Std Err	0.0462	0.0460	0.0499	0.0465	0.0461	0.0505
		RE	0.5862	0.8017	0.6357	0.5890	0.8035	0.6433
250	12	Bias	0.0000	0.0004	0.0004	0.0005	0.0007	0.0007
		Std Err	0.0261	0.0240	0.0243	0.0258	0.0240	0.0246
		RE	0.6513	0.7587	0.6601	0.6422	0.7603	0.6691
500	2	Bias	0.0015	0.0020	0.0030	0.0019	0.0020	0.0028
		Std Err	0.0613	0.0595	0.0615	0.0614	0.0596	0.0614
		RE	0.6974	0.7968	0.6884	0.6987	0.7976	0.6880
500	4	Bias	-0.0016	-0.0010	-0.0003	-0.0010	-0.0014	-0.0012
		Std Err	0.0362	0.0364	0.0370	0.0354	0.0360	0.0362
		RE	0.6880	0.8087	0.7128	0.6727	0.8014	0.6972
500	12	Bias	0.0003	0.0002	0.0003	0.0002	0.0000	0.0000
		Std Err	0.0180	0.0719	0.0184	0.0184	0.0181	0.0188
		RE	0.6096	0.7927	0.6630	0.6254	0.8015	0.6804

TABLE 4.3. *Feasible PCQRl estimation. The table considers λ^* estimated using asymptotic and bootstrapped variance. RE stands for the relative efficiency of Abrevaya and Dahl's estimator (PCQ) relative to the penalized approach (PCQRl).*

As expected, the performance of the methods that ignores the correlation between the independent variable and the individual effect are rather unsatisfactory. In all the variants of the model, the bias is significant even for moderate T. The PCQR estimator, however, reduces the variance of Abrevaya and Dahl estimator by 23 percent on average. We also see that the PCQR is more efficient than the GLS when a_i and u_{it} are drawn from t-student distribution. Keeping the design of the experiment the same, we also evaluated the performance of the estimator for the location-scale shift model assuming $\delta = .1$. The results were similar to the results in Table 4.1, revealing that the PCQR is unbiased and reduces the variance of unbiased estimators for the correlated random effects model. The results are available upon request.

We now investigate the performance of $\hat{\lambda}$ using the same models. We increase the design to consider two ways of estimating λ^*: (a) estimated asymptotic covariance matrix, and (b) bootstrapped variance. Tables 4.2 and 4.3 are different than Table 4.1 in two aspects. First, they report the performance of the PCQR at different quantiles $\{0.25, 0.5, 0.75\}$. Second, they present a measure of the efficiency of the CQR estimator relative to the PCQR estimator,

$$\mathrm{RE}^2 = \frac{\mathrm{Var}\left(\hat{\beta}_{1,PCQR}\left(\tau_j, \hat{\lambda}\right)\right)}{\mathrm{Var}\left(\hat{\beta}_{1,CQR}\left(\tau_j\right)\right)}.$$

The table presents three interesting new findings. First, the results suggest that there are no important efficiency losses when the researcher estimates λ^*, at least in the models considered in this study. Second, the performance of the two λ selection alternatives are satisfactory, and the estimation strategies seem to complement each other. Lastly, the penalized estimator seems to advance the estimator proposed by Abrevaya and Dahl (2008). The shrinkage estimator offers considerable efficiency gains over the quantile regression estimator for the correlated random effects model in all variants of the model.

5. Three Simple Examples

In this section, we use data to investigate the performance of the method, considering applications of the correlated random effects model. The first example uses a subsample of genetically identical twins from Ashenfelter and Krueger (1994) to estimate the return of education. In the second application, we investigate the distributional effect of background risk on wealth. Lastly, we estimate the intertemporal substitution elasticity of labor-supply using the British Household Panel Survey (BHPS) and considering MaCurdy (1981) and Jakubson's (1988) framework for empirical analysis. Our objective is to demonstrate how the penalized quantile regression estimator for models with endogenous individual effects can be obtained and employed.

Variable	Method	Quantiles				
		0.10	0.25	0.50	0.75	0.90
Years of Education	QR	0.060	0.087	0.095	0.091	0.083
		(0.036)	(0.030)	(0.020)	(0.015)	(0.024)
	CQR	0.058	0.097	0.101	0.042	0.123
		(0.062)	(0.045)	(0.038)	(0.045)	(0.052)
	PCQR	0.099	0.103	0.092	0.097	0.094
		(0.040)	(0.035)	(0.031)	(0.029)	(0.031)
Male Dummy	QR	0.176	0.157	0.213	0.197	0.151
		(0.151)	(0.114)	(0.074)	(0.098)	(0.177)
	CQR	0.152	0.157	0.178	0.164	0.123
		(0.173)	(0.117)	(0.071)	(0.094)	(0.156)
	PCQR	0.118	0.182	0.166	0.205	0.159
		(0.135)	(0.117)	(0.076)	(0.098)	(0.128)
White Dummy	QR	-0.326	-0.500	-0.366	-0.625	-0.464
		(0.339)	(0.187)	(0.157)	(0.198)	(0.214)
	CQR	-0.361	-0.383	-0.421	-0.495	-0.560
		(0.317)	(0.201)	(0.143)	(0.180)	(0.216)
	PCQR	-0.125	-0.393	-0.369	-0.517	-0.723
		(0.313)	(0.190)	(0.151)	(0.189)	(0.200)
Age	QR	0.070	0.081	0.105	0.087	0.104
		(0.045)	(0.029)	(0.019)	(0.026)	(0.064)
	CQR	0.065	0.071	0.101	0.101	0.095
		(0.047)	(0.029)	(0.018)	(0.027)	(0.061)
	PCQR	0.101	0.081	0.107	0.094	0.098
		(0.043)	(0.031)	(0.020)	(0.026)	(0.047)
Age^2 $\times 100$	QR	-0.068	-0.081	-0.108	-0.087	-0.111
		(0.056)	(0.036)	(0.024)	(0.030)	(0.082)
	CQR	-0.063	-0.069	-0.102	-0.102	-0.093
		(0.058)	(0.037)	(0.022)	(0.031)	(0.079)
	PCQR	-0.105	-0.083	-0.107	-0.093	-0.097
		(0.055)	(0.039)	(0.024)	(0.031)	(0.056)

TABLE 5.1. *Quantile regression estimates for the return to education model using data on Twins. The table shows results from quantile regression (QR), Abrevaya and Dahl estimator (CQR) and penalized quantile regression for the correlated random effects model (PCQR). Standard errors (in parenthesis) obtained after 1000 panel-bootstrap repetitions.*

5.1. **Example 1: Returns to Education.** Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) use a sample of genetically identical twins to investigate the return to education. Their conceptual framework includes a wage equation for the first and second twins in the i-th pair and a general representation of the individual specific effects as correlated random effects. Consider the Ashenfelter and Krueger (1994) set up,

$$\begin{aligned} y_{i1} &= \boldsymbol{x}_i'\boldsymbol{\pi} + \beta z_{i1} + \alpha_i + u_{i1} \\ y_{i2} &= \boldsymbol{x}_i'\boldsymbol{\pi} + \beta z_{i2} + \alpha_i + u_{i2} \end{aligned}$$

where y_{ij} is the logarithm of wages for twins in the j-th pair, $\boldsymbol{x}_i$ is a vector of variables that vary by families (e.g., age, gender and race) and z_{ij} represents twins characteristics (e.g., education). They consider a general representation for the individual effects as,

$$\alpha_i = \boldsymbol{x}_i'\boldsymbol{\theta} + \gamma_1 z_{i1} + \gamma_2 z_{i2} + a_i,$$

where the γ's represent the effect of education on wages that is attributed to the family effect. (While Ashenfelter and Krueger (1994) assume that $\gamma_1 = \gamma_2$, Ashenfelter and Zimmerman (1997) consider that γ_1 may be different than γ_2). It is assumed that a_i is uncorrelated with $\boldsymbol{x}_i$, z_{i1}, and z_{i2}. The conceptual framework suggest the following quantile regression model,

$$Q_{Y_{ij}}(\tau|\boldsymbol{x}_{ij}, \boldsymbol{z}_i, a_i) = \boldsymbol{x}_i'\boldsymbol{\delta}(\tau) + \beta(\tau) z_{ij} + \boldsymbol{z}_i'\boldsymbol{\gamma}(\tau) + a_i.$$

Table 5.1 presents the results obtained using the penalized method. The PCQR estimates of the return to education vary between 9 and 10 percent across the quantiles of the conditional distribution of log of wages, although QR and PQR suggest a wider range from 4 to 10 percent. We see that the standard errors of the PCQR are in general smaller than the standard errors of the CQR. This is particularly important on the lower tail, where the only significant quadratic effect on age is related to the penalized approach.

5.2. **Example 2: Distributional effects of Uncertainty on Wealth.** This section employs the framework developed by Carroll and Samwick (1998) to investigate the predictions of "buffer-stock savings" theories. In situations where the households cannot perfectly smooth their consumption, they would like to accumulate wealth to be used in the event of an income shock. These theories imply that the households will set a target wealth to permanent income ratio trying to maintain that ratio. However, the empirical evidence on rich and poor families seems to contradict this prediction. Ziliak (2003) investigates this conjecture assuming that some regressors are correlated with family specific effects α_i and classifying the families as poor, near poor, and rich. The quantile regression analysis presented below is similar and has the advantage that there is no need to create a classification according to wealth levels. It seems natural then to estimate a quantile regression model for wealth,

$$Q_{\log(W_{it})}(\tau|\boldsymbol{x}_{it}, \boldsymbol{z}_i, a_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \boldsymbol{z}_i'\boldsymbol{\delta}(\tau) + \boldsymbol{x}_i'\boldsymbol{\gamma} + a_i,$$

where $\boldsymbol{x}_{it}$ is a vector of variables (e.g., age, marital status, gender) that are correlated with the individual effect α_i. The vector $\boldsymbol{z}_i$ includes measures of permanent labor income and income uncertainty that are defined in Ziliak (2003). Note that these variables are time-invariant, therefore the within transformation cannot be used to consistently estimate the parameter $\boldsymbol{\delta}(\tau)$.

We use wealth information from 1984, 1989, 1994, 1999, and 2001 supplements of the Panel of Income Dynamics (PSID) to consider two alternative definitions for the dependent

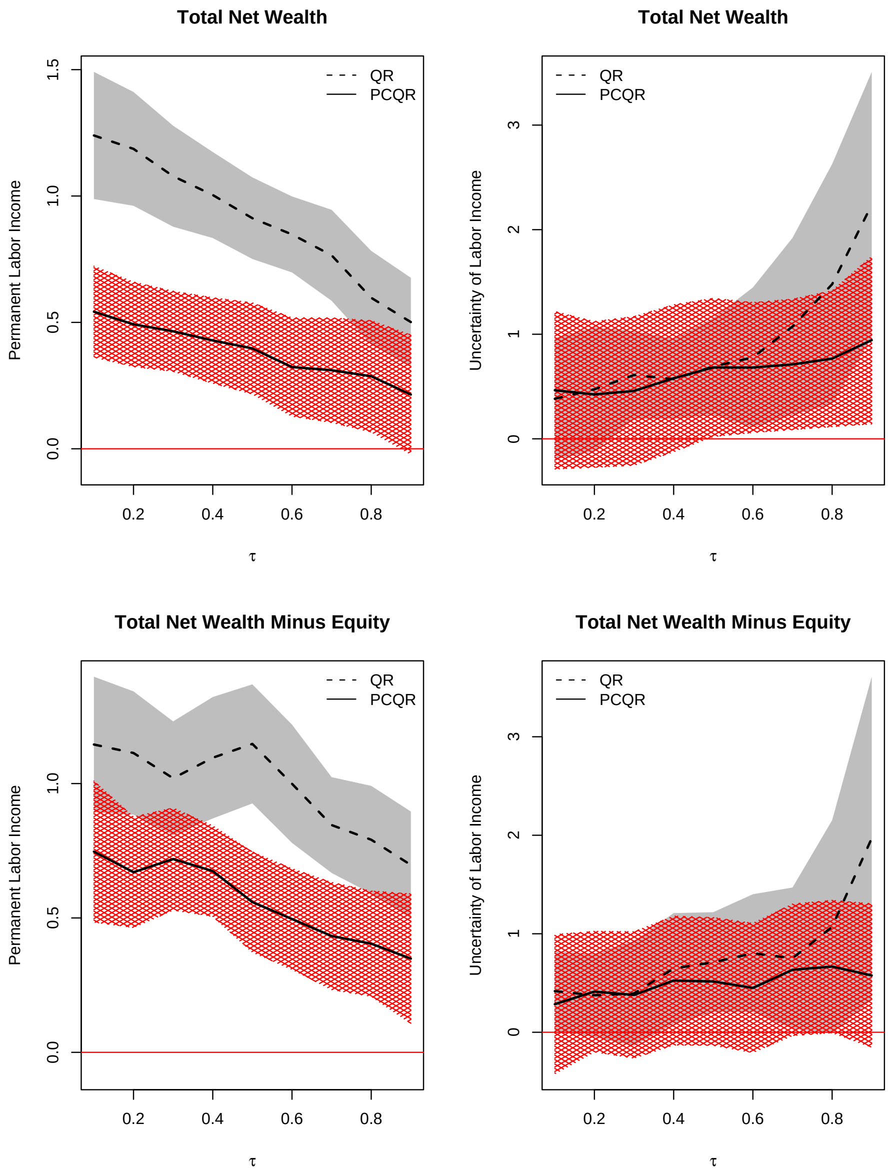

FIGURE 5.1. *The distributional effect of uncertainty and permanent labor income on wealth. The panels show quantile regression results (QR) and penalized quantile regression for the correlated effects model (PCQR). The shaded areas indicate a .95 pointwise confidence interval.*

variable W_{it}: Total Net Wealth (TNW) and Net Wealth excluding equity of the main home and personally owned business equity (NWNH). TNW is defined as the sum of house value, business equity, cash, stocks, vehicles, and other assets minus mortgage and financial debt.

The data contains 261 individuals observed over time with additional data on age, number of children, marital status, gender, and labor income.

Figure 5.1 presents estimates of the effect of uncertainty and permanent labor income as a function of the quantile τ of the conditional distribution of wealth. The upper panels show results for TNW, and the lower panels depict results for NWNH. In each graph, the continuous line denotes the estimates from the penalized approach (PCQR), and the dashed line shows the quantile regression results (QR). While QR tends to overestimate the effects, the PCQR estimates are similar to Ziliak's findings. Additionally, the median PCQR estimates are similar to the mean IV results presented in Carroll and Samwick (1998). The PCQR estimates that the effect of permanent income is positive, significant, and decreasing in terms of quantiles, and uncertainty of labor income seems to play a small role on accumulation among rich families.

5.3. **Example 3: British Evidence of Hours of Work and Wages.** Our last example considers the classical life-cycle model of consumption and labor supply assuming the following convenient additively separable utility function on consumption c and hours of work h, $c_t^{\nu_1} - s_t h_t^{\nu_2}$, where $0 < \nu_1 < 1$, $\nu_2 > 1$, and s is a taste shifter. The consumer's problem is to maximize a lifetime utility function subject to an intertemporal budget constraint. Assuming that the marginal utility of wealth is constant and that the interior optimum exists, we have

$$\ln(h_{it}) = \alpha_i + \delta \ln(w_{it}) + \gamma t - \delta \ln(s_{it}), \tag{5.1}$$

where ln denotes natural logarithm, $\delta = (\nu_2 - 1)^{-1}$ is the intertemporal substitution elasticity, and α_i represents the marginal utility of wealth that may be correlated with the independent variables. MaCurdy (1981) explicitly modeled individual specific effect as a linear function of wages, individual characteristics, and initial wealth, and more recently, Jakubson (1988) assumes the "correlated random effects" model formalizing the idea that the time invariant effect α, and the independent variables $\boldsymbol{x}$ are correlated,

$$\alpha_i = \boldsymbol{x}'_{i1}\xi_1 + ... + \boldsymbol{x}'_{iT}\xi_T + a_i.$$

where the vector $\boldsymbol{x}'_{it}$ includes log of wages and taste shifters. The quantile regression function for this model can be written as,

$$Q_{\ln(h_{it})}(\tau | \ln(w_{it}), \boldsymbol{x}_{it}, \boldsymbol{x}_i, a_i) = \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau) + \delta(\tau)\ln(w_{it}) + \boldsymbol{x}'_i\xi(\tau) + a_i.$$

We use a sample taken from the British Household Panel Survey. The data is an annual panel survey that includes 3630 observations over ten years: 1991-2000. The sample, which is similar to other data used in previous labor supply studies (e.g., PSID), includes 363 men aged between 25 and 55. The data set includes observations on weekly hours worked (mean = 46.05 and s.d. = 9.40), age (mean = 40.43 and s.d. = 6.58), and number of children

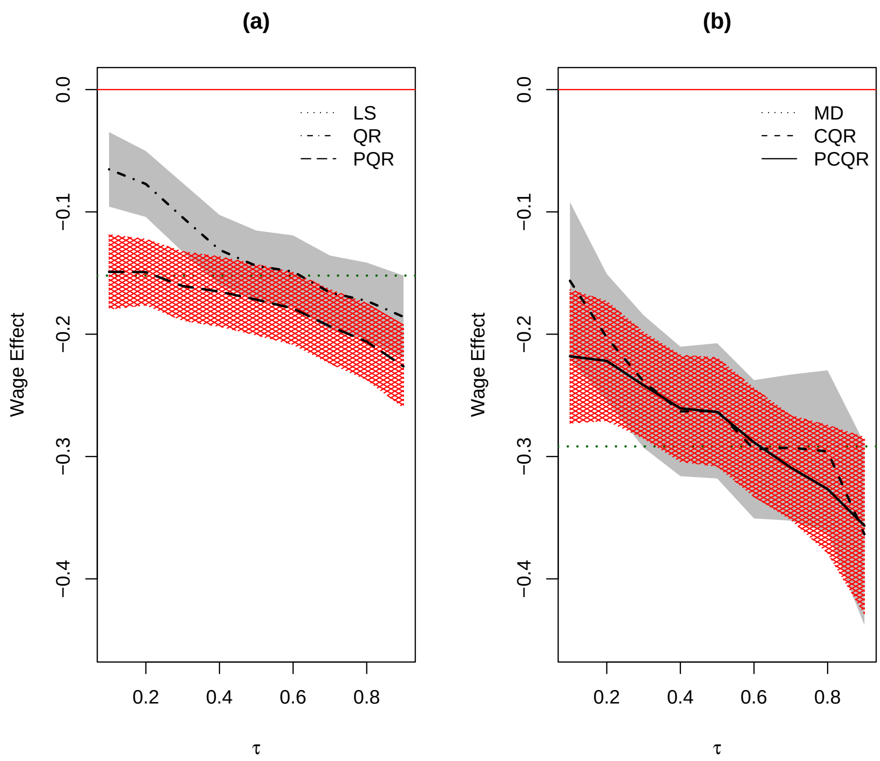

FIGURE 5.2. *The responsiveness of hours to wages using British data. The panel shows estimates obtained from least squares (LS), a version of Chamberlain's correlated random effects estimator (MD), quantile regression (QR), quantile regression for the correlated random effects model (CQR), and penalized quantile regression for the correlated random effects model (PCQR). The shaded areas indicate a .95 pointwise confidence interval.*

(mean = 1.24 and s.d. = 1.05). The British panel does not report separate information on basic and overtime earnings; therefore, we constructed the hourly gross wage in 2002 pounds considering basic and overtime hours as described in Stewart and Swaffield (1997). The logarithm of wages has a mean of 2.47 (s.d. = 44).

Figure 5.2 presents estimates of the elasticity δ as a function of the quantiles of the conditional distribution of hours. While panel (a) shows quantile regression (QR) and penalized quantile regression (PQR) results, panel (b) presents results considering the correlated random effects method (CQR) proposed by Abrevaya and Dahl (2008) and the penalized approach for the random effects (PCQR). The evidence suggests a negative elasticity of substitution for British men across the quantiles with a tendency to decrease as τ increases. The substitution effect is more pronounced at the upper tail of the conditional distribution of hours, suggesting that full-time workers value more leisure than part-time workers. We see in panel (b) the advantage of regularizing the individual effects; the estimated elasticities are similar but the PCQR gives narrower confidence intervals.

6. Conclusions and Extensions

This paper proposes a quantile approach for panel data models with endogenous individual effects. Specifically, we investigate a penalized quantile regression estimation approach relaxing the assumption that the individual effects are drawn from zero median distribution functions. We explicitly consider individual heterogeneity associated with the covariates assuming that the individual specific effects are drawn from distribution functions with location equal to a linear combination of the independent variables. This case can be motivated by the correlated random-effects model (Chamberlain 1982). We provide conditions under which the estimator is asymptotically unbiased and Gaussian, thus the harshness of the penalization can be determined by minimizing estimated variance. Small and large sample evidence reveal that the penalized approach can eliminate bias arising in models with endogenous individual effects, and significantly increase the precision relative to existing estimation methods.

Two generalizations of the approach are being investigated. First, the large sample performance of the method in the case of λ selected via bootstrap, which fits within Knight and Fu (2000) framework. Additionally, we have been developing an instrumental variable quantile method for panel data based on Chernozhukov and Hansen (2005). The approach may be needed in cases when the regressors are not only correlated with the individual effects but also with the error term. These directions appear as critical steps forward for further development of quantile regression methods for panel data.

Appendix A. Proofs

Proof of Theorem 1: The proof is completed in several steps. First, we overcome the difficulty associated to infinite dimensional vectors by concentrating out the effects into the objective function (see, e.g., Koenker 2004). In our setting, the strategy provides a convenient analytic framework for analyzing the asymptotic behavior of the penalty function. Second, we show that under the regularity conditions and the conditions over T and N, the remainder term is asymptotically negligible. Lastly, we obtain the limiting form of the objective function. Let,

$$\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i=1}^{N}\omega_j\rho_{\tau_j}(y_{it}-\xi_{it}(\tau_j)-\delta_{0i}/\sqrt{T}-\boldsymbol{x}_{it}'\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}-\boldsymbol{x}_i'\boldsymbol{\delta}_2(\tau_j)/\sqrt{N})$$
$$-\rho_{\tau_j}(y_{it}-\xi_{it}(\tau_j))+\lambda\sum_{i=1}^{N}\rho_{\tau_m}(a_i+\delta_{0i}/\sqrt{T})-\rho_{\tau_m}(a_i)$$

where τ_m is the median quantile and $\xi_{it}(\tau_j) = \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau_j) + \boldsymbol{x}'_i\boldsymbol{\gamma}(\tau_j) + a_i$ is the conditional quantile function. Following Koenker (2004), for any $(\Delta_{0i}, \Delta_1, \Delta_2) > 0$,

$$\sup_{|\delta_{0i}|<\Delta_0, ||\delta_1||<\Delta_1, ||\delta_2||<\Delta_2} ||k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - k(0, \mathbf{0}, \mathbf{0}) - \mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) - k(0, \mathbf{0}, \mathbf{0}))|| = o_p(1)$$

where

$$k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = -\frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i \psi_{\tau_j} \left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}'_i \frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j) \right)$$

with $\psi_{\tau_j}(u) = \tau_j - I(u < 0)$. Taking expectation and expanding $k(\cdot)$, we obtain

$$\mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2)) = -\frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j f_{it}(\xi_{it}(\tau_j)) \boldsymbol{x}_i \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}'_i \frac{\boldsymbol{\delta}_2(\tau_j)}{\sqrt{N}} \right) + o_p(1)$$

Optimality of the $\hat{\delta}_{2t}$'s implies that $k(\delta_{0i}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2) = o(N^{-1})$, then

$$\frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} = -\boldsymbol{h}_{jit}^{-1} \left[\sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \check{\boldsymbol{x}}_{ij} \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} \right) + \sum_{j=1}^{J} \sum_{i=1}^{N} \omega_j \boldsymbol{x}_i \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \right] + \frac{\boldsymbol{R}_N}{\sqrt{N}}$$

where $\boldsymbol{h}_{jit} = \sum_j \sum_i \omega_j \boldsymbol{x}_i \check{\boldsymbol{x}}'_{ij}$ and $\check{\boldsymbol{x}}_{ij} = (f_{it}(\xi_{it}(\tau_j))\boldsymbol{x}_{it})_{itj}$. Substituting the $\hat{\delta}_{2t}$'s, we denote

$$k(\delta_{0i}, \boldsymbol{\delta}_1) = -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j \psi_{\tau_j} \left(y_{it} - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} - \boldsymbol{x}'_i \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} - \xi_{it}(\tau_j) \right) + \frac{\lambda_T}{\sqrt{T}} \psi_{\tau_m} \left(a_i + \frac{\delta_{0i}}{\sqrt{T}} \right)$$

Again, uniformly for $|\delta_{0i}| < \Delta_{0i}$, and $||\delta_1|| < \Delta_1$, one can show that

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

Expanding as above, we obtain

$$\mathbb{E}(k(\delta_{0i}, \boldsymbol{\delta}_1)) = -\frac{1}{\sqrt{T}} \sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + \boldsymbol{x}'_i \frac{\hat{\boldsymbol{\delta}}_2(\tau_j)}{\sqrt{N}} \right) - \frac{\lambda_T}{\sqrt{T}} g_i(0) \frac{\delta_{0i}}{\sqrt{T}} + o_p(1)$$

Letting f_i denote the weighted sum of the densities for subject i, $\mu_{jit} = 1 - \boldsymbol{x}'_i \boldsymbol{h}_{jit}^{-1} \check{\boldsymbol{x}}_{ij}$, and $w_{itj} = \mu_{itj} + f_i^{-1} \lambda_T / \sqrt{T} g_i(0)$, the asymptotic (Bahadur) representation of the individual specific effect relates to the slope parameter in the following way,

$$\begin{aligned} \frac{\hat{\delta}_{0i}}{\sqrt{T}} &= -\left(\sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) w_{itj} \right)^{-1} \left(\sum_{j=1}^{J} \sum_{t=1}^{T} \omega_j f_{it}(\xi_{it}(\tau_j)) \mu_{itj} \boldsymbol{x}'_{it} \right) \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + m(\psi_{\tau_j}, R) \\ &= -\sum_{j=1}^{J} \tilde{\boldsymbol{x}}_{it}(\tau_j)' \frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}} + m(\psi_{\tau_j}, R) \end{aligned}$$

where $m(\cdot)$ is a linear function whose components are, under the regularity conditions, asymptotically negligible. Without loss of generality, we evaluate the contribution of the remainder term R

considering one quantile. The remainder term can be represented by dominant components from the Bahadur representation of $\boldsymbol{\delta}_2$ and δ_{0i},

$$R_{TN} \approx \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} f_{it} \boldsymbol{x}_{it} \left(\frac{R_{Ti}}{\sqrt{T}} + \frac{R_{Nt}}{\sqrt{N}} \right)$$

As shown in Koenker (2004), the contribution of the first term is negligible. The analysis of Knight (2001) shows that $N^{1/4} R_{Nt}$ converges in distribution to a functional of Brownian motion, therefore we write the second term as,

$$\frac{1}{N^{1/4}} \frac{K}{\sqrt{T}} \sum_{t=1}^{T} R_{0t},$$

for a generic constant K. Under the regularity conditions and the condition on the growth of T, the contribution of the remainder term is asymptotically negligible.

We replace the asymptotic representation of the individual specific effect in the objective function, and decompose the equation in four terms defined as,

$$\begin{aligned}
V_{TN}^{(1)}(\boldsymbol{\delta}_1) &= -\sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')(\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}) \psi_{\tau_j}(y_{it} - \xi_{it}(\tau_j)) \\
V_{TN}^{(2)}(\boldsymbol{\delta}_1) &= \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j \int_0^{v_{itj,TN}} (I(y_{it} - \xi_{it}(\tau_j) \leq s) - I(y_{it} - \xi_{it}(\tau_j) \leq 0)) ds \\
V_{TN}^{(3)}(\boldsymbol{\delta}_1) &= -\lambda_T \sum_{j=1}^{J} \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i(\tau_j)' (\boldsymbol{\delta}_1(\tau_j)/\sqrt{NT}) \psi_{\tau_m}(a_i) \\
V_{TN}^{(4)}(\boldsymbol{\delta}_1) &= \lambda_T \sum_{i=1}^{N} \int_0^{\tilde{\boldsymbol{x}}_i' \boldsymbol{\delta}_1/\sqrt{TN}} (I(a_i \leq s) - I(a_i \leq 0)) ds
\end{aligned}$$

with $v_{itj,TN} = (\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}$. The first term is asymptotically Gaussian. By the Lindeberg-Feller Central Limit Theorem, and conditions A3-4,

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

The second term converges in probability to a quadratic term in $\boldsymbol{\delta}_1$,

$$\mathbb{E} V_{TN}^{(2)}(\boldsymbol{\delta}_1) = \frac{1}{2TN} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N} \omega_j f_{it}(\xi_{it}(\tau_j))((\boldsymbol{x}_{it}' - \tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j))^2 + o(1) \to \frac{1}{2} \boldsymbol{\delta}_1' \boldsymbol{H}_1 \boldsymbol{\delta}_1$$

The variance of $V_{TN}^{(2)}(\boldsymbol{\delta}_1)$ converges to zero by condition A4. Similarly, by the Lindeberg-Feller Central Limit Theorem, the Slutsky Theorem, and conditions A3-4, the third term is asymptotically Gaussian,

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1) = -\frac{\lambda_T}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{j=1}^{J} \sum_{i=1}^{N} \tilde{\boldsymbol{x}}_i(\tau_j)' \boldsymbol{\delta}_1(\tau_j) \psi_{\tau_m}(a_i) \rightsquigarrow -\lambda \boldsymbol{\delta}_1' \boldsymbol{C},$$

where $\boldsymbol{C}$ is a Gaussian vector independent of $\boldsymbol{B}$ with covariance $\boldsymbol{H}_2$. The last last term has a quadratic contribution,

$$\mathbb{E}\left(V_{TN}^{(4)}(\boldsymbol{\delta}_1)\right) = \frac{\lambda_T}{2TN}\sum_{i=1}^{N} g_i(0)(\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1)^2 + o(1) \to \frac{1}{2}\lambda\boldsymbol{\delta}_1'\boldsymbol{H}_3\boldsymbol{\delta}_1$$

The proof follows since $V_{TN}(\boldsymbol{\delta}_1)$ is convex and $V_0(\boldsymbol{\delta}_1)$ has a unique minimum. ■

Proof of Theorem 2: This proof follows closely the argument developed in Koenker (2004) and Lamarche (2006). Let the location of the distribution of α_i be $s_i = \boldsymbol{x}_i'\boldsymbol{\gamma}$. Without loss of generality, we consider the location s_i in a neighborhood of 0. Let,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \xi_{it}(\tau) - \frac{\delta_{0i}}{\sqrt{T}} - \boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{NT}}\right) - \rho_\tau(y_{it} - \xi_{it}(\tau)) + \lambda_T\sum_{i=1}^{N}\rho_\tau\left(\alpha_i + \frac{\delta_{0i}}{\sqrt{T}}\right) - \rho_\tau(\alpha_i)$$

where τ is the median quantile and $\xi_{it}(\tau) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \alpha_i$ is the conditional quantile function. For any $(\Delta_{0i}, \Delta_1) > 0$,

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

where

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

with $\psi_\tau(u) = \tau - I(u < 0)$. Taking expectation and expanding v_i, we obtain

$$\begin{aligned}
\mathbb{E}v_i &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(F_{it}\left(\xi_{it}(\tau) + \frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - \tau\right) + \frac{\lambda_T}{\sqrt{T}}\left(\tau - G_i\left(-\frac{\delta_{0i}}{\sqrt{T}}\right)\right) \\
&= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_{it}(\xi_{it}(\tau))\left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - \frac{\lambda_T}{\sqrt{T}}g_i(s_i)\frac{\delta_{0i}}{\sqrt{T}} + o(1)
\end{aligned}$$

Letting $f_i = \sum_{t=1}^{T} f_{it}(\xi_{it}(\tau)) + \lambda_T g_i(s_i)$, we find that

$$\begin{aligned}
\frac{\hat{\delta}_{0i}}{\sqrt{T}} &= -\left(\sum_{t=1}^{T} f_i^{-1} f_{it}(\xi_{it}(\tau))\boldsymbol{x}_{it}'\right)\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}} + m(\psi_\tau, R) \\
&= -\dot{\boldsymbol{x}}_i'\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}} + m(\psi_\tau, R)
\end{aligned}$$

By similar conditions to the ones described in Theorem 1 and in Koenker's (2004) Theorem 1, the components of m are asymptotically negligible. Therefore we write the objective function $V_{TN}(\boldsymbol{\delta}_1(\tau))$ as,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \xi_{it}(\tau) - (\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - \rho_\tau(y_{it} - \xi_{it}(\tau)) + \lambda_T\sum_{i=1}^{N}\rho_\tau\left(\alpha_i - \frac{\dot{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - \rho_\tau(\alpha_i)$$

We now decompose the objective function in four parts:

$$\begin{aligned} V_{TN}^{(1)}(\boldsymbol{\delta}_1(\tau)) &= -\sum_{t=1}^{T}\sum_{i=1}^{N}(\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')(\boldsymbol{\delta}_1(\tau)/\sqrt{NT})\psi_\tau(y_{it} - \xi_{it}(\tau)) \\ V_{TN}^{(2)}(\boldsymbol{\delta}_1(\tau)) &= \sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{v_{it,TN}} (I(y_{it} - \xi_{it}(\tau) \leq s) - I(y_{it} - \xi_{it}(\tau) \leq 0))ds \\ V_{TN}^{(3)}(\boldsymbol{\delta}_1(\tau)) &= -\lambda_T \sum_{i=1}^{N} \dot{\boldsymbol{x}}_i' \left(\boldsymbol{\delta}_1(\tau)/\sqrt{NT}\right) \operatorname{sgn}(\alpha_i) \\ V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau)) &= \lambda_T \sum_{i=1}^{N} \int_0^{\dot{\boldsymbol{x}}_i' \frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}} (I(\alpha_i \leq s) - I(\alpha_i \leq 0))ds \end{aligned}$$

with $v_{it,TN} = (\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')\boldsymbol{\delta}_1(\tau)/\sqrt{TN}$. The first two parts corresponds to the decomposition of the check function $\rho_\tau(\cdot)$, and the last two parts corresponds to the decomposition of the penalty term $P(\cdot)$. The first term is asymptotically Gaussian,

$$V_{TN}^{(1)}(\boldsymbol{\delta}_1(\tau)) = -\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N}(\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')\boldsymbol{\delta}_1(\tau)\psi_\tau(y_{it} - \xi_{it}(\tau)) \rightsquigarrow -\boldsymbol{\delta}_1(\tau)'\boldsymbol{B}$$

where $\boldsymbol{B}$ is a Gaussian vector with covariance $\boldsymbol{H}_0$. The second term converges in probability to a quadratic term in $\boldsymbol{\delta}_1(\tau)$. Note that

$$\mathbb{E}V_{TN}^{(2)}(\boldsymbol{\delta}_1(\tau)) = \frac{1}{2TN}\sum_{t=1}^{T}\sum_{i=1}^{N} f_{it}(\xi_{it}(\tau))((\boldsymbol{x}_{it}' - \dot{\boldsymbol{x}}_i')\boldsymbol{H}_1(\tau))^2 + o(1) \to \frac{1}{2}\boldsymbol{\delta}_1(\tau)'\boldsymbol{H}_1\boldsymbol{\delta}_1(\tau)$$

The last two terms of $V_{TN}(\boldsymbol{\delta}_1(\tau))$ represents a decomposition of the stochastic penalty term. The third term is also asymptotically Gaussian,

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1(\tau)) = -\frac{\lambda_T}{\sqrt{T}}\boldsymbol{\delta}_1(\tau)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\dot{\boldsymbol{x}}_i'\operatorname{sgn}(\alpha_i) \rightsquigarrow -\lambda\boldsymbol{\delta}_1(\tau)'\boldsymbol{C}$$

where $\boldsymbol{C}$ is a Gaussian vector independent of $\boldsymbol{B}$ with covariance $\boldsymbol{H}_2$. Lastly, the fourth term $V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau))$ is asymptotically quadratic in $\boldsymbol{\delta}_1(\tau)$,

$$\mathbb{E}V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau)) = \frac{\lambda_T}{2TN}\sum_{i=1}^{N} g_i(s_i)(\dot{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau))^2 + o(1) \to \frac{1}{2}\lambda\boldsymbol{\delta}_1(\tau)'\boldsymbol{H}_3\boldsymbol{\delta}_1(\tau)$$

Since $V_{TN}(\boldsymbol{\delta}_1(\tau))$ is convex, and $V_0(\boldsymbol{\delta}_1(\tau))$ has a unique minimum, it follows that

$$\operatorname{argmin}(V_{TN}(\boldsymbol{\delta}_1(\tau))) = \sqrt{TN}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \operatorname{argmin}(V_0(\boldsymbol{\delta}_1(\tau)))$$

Noting that the limiting form of the objective function is then,

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

we find that its minimizer is $\boldsymbol{v} = \boldsymbol{\delta}_1(\tau) = (\boldsymbol{H}_1 + \lambda\boldsymbol{H}_3)^{-1}(\boldsymbol{B} + \lambda\boldsymbol{C})$. The proof is completed in Section 3.2. ∎

Proof of Remark 3: We first obtain the expression for the estimator that penalizes endogenous individual effects. The trace of the asymptotic covariance matrix of the penalized estimator $\tilde{\boldsymbol{\beta}}(\tau,\lambda)$ can be written as

$$\text{trAvar}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) \quad = \quad \text{tr}\left\{(\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{C}}(\tilde{\boldsymbol{D}}+\lambda^2\boldsymbol{I})(\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1}\right\},$$

where the matrices $\tilde{\boldsymbol{A}} = \boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Phi}}'\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}$, $\tilde{\boldsymbol{B}} = \tilde{\boldsymbol{A}}^{-1}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{\Phi}\boldsymbol{M}\boldsymbol{X}$, $\tilde{\boldsymbol{C}} = \boldsymbol{X}'\boldsymbol{P}_{\boldsymbol{\Phi}}'\boldsymbol{P}_{\boldsymbol{\Phi}}\boldsymbol{X}$, and $\tilde{\boldsymbol{D}} = \tau(1-\tau)\tilde{\boldsymbol{C}}^{-1}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X}$. Replacing the matrices by their spectral decomposition, the trace of the asymptotic covariance matrix is,

$$\text{tr}\{(\boldsymbol{U}_{\tilde{b}}\boldsymbol{\Lambda}_{\tilde{b}}\boldsymbol{U}_{\tilde{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\tilde{a}}\boldsymbol{\Lambda}_{\tilde{a}}\boldsymbol{U}_{\tilde{a}}')^{-1}\boldsymbol{U}_{\tilde{c}}\boldsymbol{\Lambda}_{\tilde{c}}\boldsymbol{U}_{\tilde{c}}'(\boldsymbol{U}_{\tilde{d}}\boldsymbol{\Lambda}_{\tilde{d}}\boldsymbol{U}_{\tilde{d}}'+\lambda^2\boldsymbol{I})(\boldsymbol{U}_{\tilde{b}}\boldsymbol{\Lambda}_{\tilde{b}}\boldsymbol{U}_{\tilde{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\tilde{a}}\boldsymbol{\Lambda}_{\tilde{a}}\boldsymbol{U}_{\tilde{a}}')^{-1}\},$$

or alternatively,

$$\text{tr}\{\boldsymbol{U}_{\tilde{b}}'(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_{\tilde{b}}\boldsymbol{U}_{\tilde{a}}'\boldsymbol{\Lambda}_{\tilde{a}}^{-1}\boldsymbol{U}_{\tilde{a}}\boldsymbol{U}_{\tilde{c}}'\boldsymbol{\Lambda}_{\tilde{c}}\boldsymbol{U}_{\tilde{c}}\boldsymbol{U}_{\tilde{d}}'(\boldsymbol{\Lambda}_{\tilde{d}}+\lambda^2\boldsymbol{I})\boldsymbol{U}_{\tilde{d}}\boldsymbol{U}_{\tilde{b}}'(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_{\tilde{b}}\boldsymbol{U}_{\tilde{a}}'\boldsymbol{\Lambda}_{\tilde{a}}^{-1}\boldsymbol{U}_{\tilde{a}}\}.$$

Since the trace of $\boldsymbol{ABA}$ is equal to the trace of $\boldsymbol{AAB}$ and $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}$, the equation is now,

$$\text{trAvar}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) \quad = \quad \text{tr}\left\{(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_{\tilde{a}})^{-1}\boldsymbol{\Lambda}_{\tilde{c}}(\boldsymbol{\Lambda}_{\tilde{d}}+\lambda^2\boldsymbol{I})(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_{\tilde{a}})^{-1}\right\}.$$

Similarly, the trace of the asymptotic bias of the penalized estimator $\tilde{\boldsymbol{\beta}}(\boldsymbol{\tau},\lambda)$ can be written as,

$$\begin{aligned}
\text{trAbias}(\tilde{\boldsymbol{\beta}}(\tau,\lambda)) &= \text{tr}\left\{\lambda^2(\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1}\boldsymbol{S}_o\boldsymbol{S}_o'((\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1})'\right\}\\
&= \text{tr}\left\{\lambda^2\boldsymbol{S}_o\boldsymbol{S}_o'(\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1}((\tilde{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\tilde{\boldsymbol{A}}^{-1})'\right\}\\
&= \text{tr}\left\{\lambda^2\boldsymbol{U}_{S_o}\boldsymbol{\Lambda}_{S_o}\boldsymbol{U}_{S_o}'(\boldsymbol{U}_{\tilde{b}}\boldsymbol{\Lambda}_{\tilde{b}}\boldsymbol{U}_{\tilde{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\tilde{a}}\boldsymbol{\Lambda}_{\tilde{a}}\boldsymbol{U}_{\tilde{a}}')^{-1}((\boldsymbol{U}_{\tilde{b}}\boldsymbol{\Lambda}_{\tilde{b}}\boldsymbol{U}_{\tilde{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\tilde{a}}\boldsymbol{\Lambda}_{\tilde{a}}\boldsymbol{U}_{\tilde{a}}')^{-1})'\right\}\\
&= \text{tr}\left\{\lambda^2\boldsymbol{\Lambda}_{S_o}(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_{\tilde{a}}^{-1}(\boldsymbol{\Lambda}_{\tilde{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_{\tilde{a}}^{-1}\right\}.
\end{aligned}$$

We now derive of the asymptotic MSE of $\hat{\boldsymbol{\beta}}(\tau,\lambda)$. The trace of the asymptotic covariance matrix can be written as

$$\text{trAMSE}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) \quad = \quad \text{tr}\left\{(\hat{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\hat{\boldsymbol{A}}^{-1}\hat{\boldsymbol{C}}(\hat{\boldsymbol{D}}+\lambda^2\boldsymbol{I})(\hat{\boldsymbol{B}}+\lambda\boldsymbol{I})^{-1}\hat{\boldsymbol{A}}^{-1}\right\},$$

where the matrices $\hat{\boldsymbol{A}} = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{\Psi}\boldsymbol{P}\boldsymbol{X}$, $\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{\Upsilon}\boldsymbol{W}\boldsymbol{X}$, $\hat{\boldsymbol{C}} = \tau(1-\tau)\boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}$, and $\hat{\boldsymbol{D}} = \tau(1-\tau)\hat{\boldsymbol{C}}^{-1}\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}$. Replacing the matrices by their spectral decomposition, the trace of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ can be written as,

$$\text{tr}\{(\boldsymbol{U}_{\hat{b}}\boldsymbol{\Lambda}_{\hat{b}}\boldsymbol{U}_{\hat{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\hat{a}}\boldsymbol{\Lambda}_{\hat{a}}\boldsymbol{U}_{\hat{a}}')^{-1}\boldsymbol{U}_{\hat{c}}\boldsymbol{\Lambda}_{\hat{c}}\boldsymbol{U}_{\hat{c}}'(\boldsymbol{U}_{\hat{d}}\boldsymbol{\Lambda}_{\hat{d}}\boldsymbol{U}_{\hat{d}}'+\lambda^2\boldsymbol{I})(\boldsymbol{U}_{\hat{b}}\boldsymbol{\Lambda}_{\hat{b}}\boldsymbol{U}_{\hat{b}}'+\lambda\boldsymbol{I})^{-1}(\boldsymbol{U}_{\hat{a}}\boldsymbol{\Lambda}_{\hat{a}}\boldsymbol{U}_{\hat{a}}')^{-1}\},$$

or,

$$\text{tr}\{\boldsymbol{U}_{\hat{b}}'(\boldsymbol{\Lambda}_{\hat{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_{\hat{b}}\boldsymbol{U}_{\hat{a}}'\boldsymbol{\Lambda}_{\hat{a}}^{-1}\boldsymbol{U}_{\hat{a}}\boldsymbol{U}_{\hat{c}}'\boldsymbol{\Lambda}_{\hat{c}}\boldsymbol{U}_{\hat{c}}\boldsymbol{U}_{\hat{d}}'(\boldsymbol{\Lambda}_{\hat{d}}+\lambda^2\boldsymbol{I})\boldsymbol{U}_{\hat{d}}\boldsymbol{U}_{\hat{b}}'(\boldsymbol{\Lambda}_{\hat{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{U}_{\hat{b}}\boldsymbol{U}_{\hat{a}}'\boldsymbol{\Lambda}_{\hat{a}}^{-1}\boldsymbol{U}_{\hat{a}}\}.$$

Again, because tr$\boldsymbol{ABA}=$ tr$\boldsymbol{AAB}$ and $\boldsymbol{U}'\boldsymbol{U}=\boldsymbol{I}$, the equation can be written now as,

$$\text{trAMSE}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) \quad = \quad \text{tr}\left\{(\boldsymbol{\Lambda}_{\hat{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_{\hat{a}}^{-1}\boldsymbol{\Lambda}_{\hat{c}}(\boldsymbol{\Lambda}_{\hat{d}}+\lambda^2\boldsymbol{I})(\boldsymbol{\Lambda}_{\hat{b}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{\Lambda}_{\hat{a}}^{-1}\right\}.$$

■

Lemma 1. *Let $\mathcal{A}=[0,\bar{\lambda})\subset\mathbb{R}_+$ where $\bar{\lambda}=1.5\zeta_f/\zeta_d+0.5\zeta_d>0$ for positive constants ζ_d and ζ_f. Then, the set $\mathcal{A}$ is non-empty, closed and bounded.*

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

Lemma 3. *Let the ζ's indicate positive constants and $\lambda \in \mathcal{A}$. Then, the rational function $\pi(\lambda)$: $\mathcal{A} \to \mathbb{R}_+$,*

$$\pi(\lambda) \quad = \quad \zeta_a^2 \zeta_e (\zeta_f + \lambda^2) / \zeta_b \zeta_c^2 (\zeta_d + \lambda)^2$$

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

Proof: Note that the function π is a rational function, and every rational function is continuous. The first and second derivative of the function $\pi(\lambda)$ with respect to λ is

$$\begin{aligned} \frac{\partial \pi(\lambda)}{\partial \lambda} &= \frac{2\zeta_a^2 \zeta_e (\zeta_d \lambda - \zeta_f)}{\zeta_b \zeta_c^2 (\zeta_d + \lambda)^3} \\ \frac{\partial^2 \pi(\lambda)}{\partial \lambda^2} &= \frac{2\zeta_a^2 \zeta_e (\zeta_d^2 - 2\zeta_d \lambda + 3\zeta_f)}{\zeta_b \zeta_c^2 (\zeta_d + \lambda)^4} > 0. \end{aligned}$$

■

Lemma 4. *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 of Theorem 3: The normalized asymptotic covariance matrix of the penalized estimator is defined as,

$$\text{AVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) \quad = \quad (\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1} (\boldsymbol{\Sigma}_0 + \lambda^2 \boldsymbol{\Sigma}_2)(\boldsymbol{\Sigma}_1 + \lambda \boldsymbol{\Sigma}_3)^{-1}$$

Letting tr denote the trace of a matrix, we write the covariance matrix as,

$$\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) \quad = \quad \text{tr}\left\{ \boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}(\boldsymbol{D} + \lambda \boldsymbol{I})^{-1} \boldsymbol{C}^{-1} \boldsymbol{E}(\boldsymbol{F} + \lambda^2 \boldsymbol{I})(\boldsymbol{D} + \lambda \boldsymbol{I})^{-1} \boldsymbol{C}^{-1} \right\},$$

where the matrices $\boldsymbol{A} = \boldsymbol{X}'\boldsymbol{W}'\boldsymbol{\Upsilon}\boldsymbol{W}\boldsymbol{X}$, $\boldsymbol{B} = \tau(1-\tau)\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X}$, $\boldsymbol{C} = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{\Psi}\boldsymbol{P}\boldsymbol{X}$, $\boldsymbol{D} = \boldsymbol{C}^{-1}\boldsymbol{A}$, $\boldsymbol{E} = \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X}$ and $\boldsymbol{F} = \boldsymbol{E}^{-1}\boldsymbol{B}$. Replacing the matrices by their spectral decomposition, the trace of the asymptotic covariance matrix can be written as,

$$\begin{aligned} \text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) &= \text{tr}\{\boldsymbol{U}_a \boldsymbol{\Lambda}_a \boldsymbol{U}_a' (\boldsymbol{U}_b \boldsymbol{\Lambda}_b \boldsymbol{U}_b')^{-1} \boldsymbol{U}_a \boldsymbol{\Lambda}_a \boldsymbol{U}_a' (\boldsymbol{U}_d \boldsymbol{\Lambda}_d \boldsymbol{U}_d' + \lambda \boldsymbol{I})^{-1} (\boldsymbol{U}_c \boldsymbol{\Lambda}_c \boldsymbol{U}_c')^{-1} \\ &\quad \boldsymbol{U}_e \boldsymbol{\Lambda}_e \boldsymbol{U}_e' (\boldsymbol{U}_f \boldsymbol{\Lambda}_f \boldsymbol{U}_f' + \lambda^2 \boldsymbol{I})(\boldsymbol{U}_d \boldsymbol{\Lambda}_d \boldsymbol{U}_d' + \lambda \boldsymbol{I})^{-1} (\boldsymbol{U}_c \boldsymbol{\Lambda}_c \boldsymbol{U}_c')^{-1}\} \\ &= \text{tr}\{\boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_b^{-1} \boldsymbol{\Lambda}_a (\boldsymbol{\Lambda}_d + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Lambda}_c^{-1} \boldsymbol{\Lambda}_e (\boldsymbol{\Lambda}_f + \lambda^2 \boldsymbol{I})(\boldsymbol{\Lambda}_d + \lambda \boldsymbol{I})^{-1} \boldsymbol{\Lambda}_c^{-1}\} \\ &= \sum_{i=1}^{p} \frac{(\zeta_a^i)^2 \zeta_e^i (\zeta_f^i + \lambda^2)}{\zeta_b^i (\zeta_c^i (\zeta_d^i + \lambda))^2} = \sum_{i=1}^{p} \pi(\lambda)^i \end{aligned}$$

We now have a simple optimization problem as a function of λ, with positive eigenvalues ζ_k^i for all i, k. By Lemma 3, π^i is convex on $\mathcal{A}^i$. Since the sets $\mathcal{A}^i$ are a decreasing sequence of sets, Lemma 4 implies that π^i is also convex on $\mathcal{D}$. Since the sum of convex functions is also convex, $\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda))$ is convex on $\mathcal{D}$. Therefore, $\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau, \lambda)) : \mathcal{D} \to \mathbb{R}_+$ is a continuous strictly convex function defined on a non-empty, compact set (Lemma 1). These sufficient conditions imply that the trace of the normalized asymptotic covariance matrix has a unique minimizer λ^* such that,

$$\text{trAVar}(\hat{\boldsymbol{\beta}}(\tau,\lambda^*)) < \text{trAVar}(\hat{\boldsymbol{\beta}}(\tau,\lambda)) \quad \forall \lambda \in \mathcal{D}$$

■

Appendix B. Computational Aspects

The quantile regression minimization problem can be formulated as a linear program,

$$\min \left\{ \tau \boldsymbol{\iota}_n' \boldsymbol{r}^+ + (1-\tau)\boldsymbol{\iota}_n' \boldsymbol{r}^- | \boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{r}^+ - \boldsymbol{r}^-, (\boldsymbol{b}, \boldsymbol{r}^+, \boldsymbol{r}^-) \in \mathcal{R}^p \times \mathcal{R}^{2n} \right\}$$

where $\boldsymbol{\iota}$ is a vector of n ones. The previous problem has a dual formulation,

$$\max \left\{ \boldsymbol{y}'\boldsymbol{d} | \boldsymbol{X}'\boldsymbol{d} = \boldsymbol{0}, \boldsymbol{d} \in [\tau - 1, \tau]^n \right\},$$

and, for $\boldsymbol{c} = \boldsymbol{d} + 1 - \tau$, may be expressed as,

$$\max \left\{ \boldsymbol{y}'\boldsymbol{c} | \boldsymbol{X}'\boldsymbol{c} = (1-\tau)\boldsymbol{X}'\boldsymbol{\iota}_n, \boldsymbol{c} \in [0,1]^n \right\}.$$

Algorithms available in the `quantreg` package of the public domain dialect `R` are based on the previous representation. See Koenker (2005, chapter 6) for a more extensive discussion. The dual representation provides a way forward for the penalized estimator considering a simple data augmentation scheme. Considering the dual and transforming variables similarly as before, we have

$$\max_{\boldsymbol{v}} \left\{ (\boldsymbol{\iota}_J \otimes \boldsymbol{y} \vdots \boldsymbol{0})'\boldsymbol{v} \mid \boldsymbol{B}'\boldsymbol{v} = \boldsymbol{h}, \boldsymbol{v} \in [0,1]^{JNT} \right\}, \tag{B.1}$$

with the vector $\boldsymbol{h} = (\boldsymbol{h}_x', \boldsymbol{h}_d', \boldsymbol{h}_z')'$ and $\boldsymbol{h}_x = (\omega_1(1-\tau_1)\boldsymbol{X}'\boldsymbol{\iota}_{NT}, \ldots, \omega_J(1-\tau_J)\boldsymbol{X}'\boldsymbol{\iota}_{NT})$, $\boldsymbol{h}_d = (\omega_1(1-\tau_1)(\boldsymbol{Z}\boldsymbol{D})'\boldsymbol{\iota}_{NT}, \ldots, \omega_J(1-\tau_J)(\boldsymbol{Z}\boldsymbol{D})'\boldsymbol{\iota}_{NT})$, and $\boldsymbol{h}_z = (\sum_j \omega_j(1-\tau_j)\boldsymbol{Z}'\boldsymbol{\iota}_{NT} + \lambda_o \boldsymbol{I}_N \boldsymbol{\iota}_N)'$. The parameter λ_o is the tuning parameter and $[0,1]^{JNT}$ denotes the JNT Cartesian product of the closed interval $[0,1]$. Obviously the design matrix is now of large dimension but the problem can be efficiently solved considering the sparse matrix algebra storage used in Koenker (2004).

All programs were written in `R`. The quantile regression method designed to estimate location shift effects $\boldsymbol{\gamma}$'s solves the dual of the problem formulated in (B.1) considering $\boldsymbol{\gamma}_j = \boldsymbol{\gamma}$ for all j. Results and programs are available upon request.

References

Abrevaya, J., and C. Dahl (2008): "The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Regression Estimation on Panel Data," conditionally accepted, Journal of Business and Economics Statistics.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002): "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experience," *American Economic Review*, 92(5), 1535–1558.

Angrist, J., E. Bettinger, and M. Kremer (2006): "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," *American Economic Review*, 96(3), 847–862.

Ashenfelter, O., and A. Krueger (1994): "Estimates of the Economic Return to Schooling from a New Sample of Twins," *American Economic Review*, 84(5), 1157–1173.

Ashenfelter, O., and C. Rouse (1998): "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins," *The Quarterly Journal of Economics*, 113(1), 253–284.

Ashenfelter, O., and D. Zimmerman (1997): "Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers," *The Review of Economics and Statistics*, 79(1), 1–9.

Bhatia, R. (1997): *Matrix Analysis*. Springer-Verlag, New York.

Bickel, P., and B. Li (2006): "Regularization in Statistics," *Sociedad de Estadistica e Investigacion Operativa*, 15.

Buchinsky, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics*, 68, 303–338.

Carey, K. (1997): "A Panel Data Design for Estimation of Hospital Cost Functions," *The Review of Economics and Statistics*, 79(3), 443–453.

Carrasco, M., J. P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics, Edited by J. Heckman and E. Leamer*, 6B, Elsevier.

Carroll, C. D., and A. A. Samwick (1998): "How Important is Precautionary Saving?," *The Review of Economics and Statistics*, 80, 410–419.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

——— (1984): "Panel Data," *Handbook of Econometrics, Edited by Z. Griliches and M.D. Intriligator*, 2, Amsterdam: North-Holland.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics, Edited by J. Heckman and E. Leamer*, 6B, Elsevier.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–262.

De Silva, D. G., G. Kosmopoulou, and C. Lamarche (2008): "The Effect of Information on the Bidding and Survival of Entrants in Procurement Auctions," *forthcoming, Journal of Public Economics*.

Geraci, M., and M. Bottai (2006): "Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution," *Biostatistics*.

Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11(1), 105–121.

Horowitz, J. (1998): "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1352.

Horowitz, J., and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75(4), 1191–1208.

Jakubson, G. (1988): "The Sensitivity of Labor-Suply Parameter Estimates to Unobserved Individual Effects: Fixed- and Random Effects Estimates in a Nonlinear Model Using Panel Data," *Journal of Labor Economics*, 6(3), 302–329.

Knight, K. (2005): "Comparing Conditional Quantile Estimators: First and Second Order Considerations," Department of Statistics, University of Toronto.

Knight, K., and W. Fu (2000): "Asymptotics for Lasso-typo estimators," *Annals of Statistics*, 28, 1356–1378.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press.

Krashinsky, H. A. (2004): "Do Marital Status and Computer Usage Really Change the Wage Structure?," *The Journal of Human Resources*, 39(3), 774–791.

Lamarche, C. (2006): "Quantile Regression for Panel Data," Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

MaCurdy, T. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003): *Semiparametric Regression.* Cambridge University Press.

Stewart, M. B., and J. K. Swaffield (1997): "Constraints on the Desired Hours of Work of British Men," *The Economic Journal*, 107, 520–535.

Tibshirani, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society B*, 58, 267–288.

Wand, M. P. (1999): "On the Optimal Amount of Shrinkage in Penalised Spline Regression," *Biometrika*, 86(4), 936–940.

Ziliak, J. P. (2003): "Income Transfers and Assets of the Poor," *The Review of Economics and Statistics*, 85(1), 63–76.