ROBUST PENALIZED QUANTILE REGRESSION ESTIMATION FOR PANEL DATA*

CARLOS LAMARCHE†

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

Keywords: Shrinkage, Robust, Quantile Regression, Panel Data, Individual Effects.

JEL Codes: C13, C23.

1. Introduction

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

Koenker (2004) suggested a quantile regression approach for panel data. He introduced a class of penalized quantile regression estimators providing a novel solution to the recognized difficulties of quantile regression for additive random effects models (Koenker and Hallock 2000, Abrevaya and Dahl 2005). The standard least squares transformations to deal with

*This Version: October 30, 2006. This paper is based on Chapter 1 of my dissertation "Quantile Regression for Panel Data" at the University of Illinois at Urbana-Champaign. I especially thank my advisor Roger Koenker for advice, insightful comments and suggestions. I am also grateful to Rick Ashley, Anil Bera, Bruce Hansen, Zhongjun Qu, Aris Spanos, and Zhijie Xiao for comments and valuable conversations, as well as seminar participants at UC Merced, University of Oklahoma, Virginia Tech, Robert Bohrer workshop at the University of Illinois at Urbana-Champaign, and the 16th Midwest Econometric Group conference. All errors are mine.

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

a large number of parameters are not available in quantile regression, so the approach proposes to estimate directly a vector of individual effects. The estimation of these (nuisance) parameters increases the variability of the estimates of the covariate effects, but regularization, or shrinkage of these effects toward a common value helps to reduce the inflation effect. An ℓ_1 penalty term (Tibshirani 1996, Donoho et al. 1998) serves to shrink the vector of individual effects, and a tuning parameter λ controls the degree of this shrinkage. Although Koenker (2004) shows that some degree of shrinkage is often desirable, finding precisely the value of λ remains unclear outside idealized Gaussian conditions. This paper investigates this issue showing that the class of penalized estimators for models with exogenous regressors are asymptotically unbiased and Gaussian. The parameter λ can thus be selected to minimize estimated asymptotic variance.

It has been acknowledged by others that the optimal choice of the regularization parameter λ is an interesting problem of both theoretical and practical importance (Hastie, Tibshirani, and Zou 2004). For instance, in model selection, the regularization parameter is selected by AIC (Akaike 1973), BIC (Schwartz 1978), in ridge regression, it is selected by minimizing mean square error (Hoerl and Kennard 1988), and in non parametrics, the choice of λ is analogous to select the smoothing parameter by cross-validation (e.g. CV, GCV). In panel data, the classical random effects approach suggests maximum likelihood (MLE) or generalized least squares (GLS) methods, but preliminary Gaussian λ selection strategies under non-classical assumptions could lead to incorrect inference. The approach considered in this paper offers a robust alternative for λ selection.

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

The next section reviews the classical theory of panel data, and presents the model and estimator. Section 3 is devoted to the asymptotic behavior of the estimator. In section 4, we obtain the optimal tuning parameter and its estimator, and in section 5, we offer Monte-Carlo evidence. In Section 6, we demonstrate how the penalized estimator can be obtained and used in empirical applications. Section 7 provides conclusions.

2. Panel Data Methods and Models

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 distributed independently across subjects, and u_{it} is an iid error term.

2.1. Classical Estimators Revisited.

It is convenient to write equation (2.1) as

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

where $\boldsymbol{Z}$ is an "incidence matrix" of dummy variables, and $\boldsymbol{\alpha}$ and $\boldsymbol{u}$ are independent random vectors. The parameter of primary interest $\boldsymbol{\beta}$ can be estimated by two alternative methods.

2.1.1. *Fixed Effects Estimator.* The fixed effects model assumes that the unobserved individual effect is an unknown parameter to be estimated (i.e., individual intercepts). The fixed effects estimator is given by the solution of the following quadratic problem

$$\min_{(\boldsymbol{\alpha},\boldsymbol{\beta})} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\|^2, \tag{2.3}$$

where $||\boldsymbol{u}||^2_{\boldsymbol{R}^{-1}}$ means $\boldsymbol{u}'\boldsymbol{R}^{-1}\boldsymbol{u}$. The classical fixed effects estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} \;=\; (\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}\boldsymbol{y}$$

where the matrices $\boldsymbol{M} \equiv \boldsymbol{I} - \boldsymbol{P}$ and $\boldsymbol{P} \equiv \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$.

2.1.2. *Random Effects Estimator.* Classical theory often alternatively assumes that $\boldsymbol{u}$ and $\boldsymbol{\alpha}$ are independent Gaussian vectors with $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, and $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{W})$. Therefore, the vector $\boldsymbol{v} = \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{u}$ has covariance matrix

$$\mathbb{E}(\boldsymbol{v}\boldsymbol{v}') = (\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}' + \boldsymbol{R}) = \boldsymbol{V}$$

The generalized least squares estimator (GLS) of $\boldsymbol{\beta}$ can be obtained either by weighting the variance components,

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2_{\boldsymbol{V}^{-1}} \tag{2.4}$$

or by estimating the unobserved specific effects from the penalized least squares problem,

$$\min_{(\boldsymbol{\alpha},\boldsymbol{\beta})} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\|^2_{\boldsymbol{R}^{-1}} + \|\boldsymbol{\alpha}\|^2_{\boldsymbol{W}^{-1}} \tag{2.5}$$

Proposition 1. $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y})$ *solves both (2.4) and (2.5).*

The estimator derived from the long objective function (2.5) is the best linear unbiased predictor (BLUP). This method for estimating $\hat{\boldsymbol{\beta}}$ and the associated random effects predictors $\hat{\boldsymbol{\alpha}}$ has been used since the 50s in statistics (see, e.g., Robinson (1991)). Under classical assumptions for the variance components (i.e., $\boldsymbol{W} = \sigma^2_\alpha I_N$ and $\boldsymbol{R} = \sigma^2_u I_{NT}$), (2.5) simplifies to

$$\min_{(\boldsymbol{\alpha},\boldsymbol{\beta})} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\|^2 + \lambda\|\boldsymbol{\alpha}\|^2, \tag{2.6}$$

where the penalty $\lambda\|\boldsymbol{\alpha}\|^2$ serves to shrink the individual effects estimates toward zero to improve the performance of the estimate of $\boldsymbol{\beta}$, and the tuning parameter $\lambda = \sigma^2_u/\sigma^2_\alpha$ controls the degree of shrinkage. The GLS estimator of $\boldsymbol{\beta}$ is,

$$\hat{\boldsymbol{\beta}}(\lambda) = (\boldsymbol{X}'\boldsymbol{V}(\lambda)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\lambda)^{-1}\boldsymbol{y},$$

where $\boldsymbol{V}(\lambda)^{-1} = [\boldsymbol{I} - (T+\lambda)^{-1}\boldsymbol{Z}\boldsymbol{Z}']$. The estimator is unbiased for all positive λ, and obviously it can be represented as a function of $\boldsymbol{V}^{-1}$, the variance matrix of the GLS estimator

$$\boldsymbol{V}^{-1} = \frac{1}{\sigma^2_u}\left[\boldsymbol{I} - \frac{\sigma^2_\alpha}{T\sigma^2_\alpha + \sigma^2_u}\boldsymbol{Z}\boldsymbol{Z}'\right]$$

This reparametrization offers an alternative view on Maddala's (1971) result, which states that fixed effects (FE), pooled least squares (OLS), and GLS estimators weight differently the between-group variation. It follows that,

$$\begin{aligned}\lim_{\lambda\to 0}\hat{\boldsymbol{\beta}}(\lambda) &= \hat{\boldsymbol{\beta}}_{FE}(0) = (\boldsymbol{X}'\boldsymbol{V}(0)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(0)^{-1}\boldsymbol{y}\\ \lim_{\lambda\to\infty}\hat{\boldsymbol{\beta}}(\lambda) &= \hat{\boldsymbol{\beta}}_{OLS}(\infty) = (\boldsymbol{X}'\boldsymbol{V}(\infty)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\infty)^{-1}\boldsymbol{y}\end{aligned}$$

2.2. **Quantile Regression Model and Estimator.** Because the error term u_{it} in (2.1) is assumed to be mean zero and orthogonal to the independent variables, the conditional mean function of the unobserved effects model is

$$\mathbb{E}(y_{it}|\boldsymbol{x}_{it},\alpha_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i,$$

where y_{it} is the response, $\boldsymbol{x}_{it}$ is the vector of covariates, and α_i is an individual fixed effect. In this paper, we consider the analogous conditional quantile model of the form

$$Q_{Y_{it}}(\tau_j|\boldsymbol{x}_{it},\alpha_i) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau_j) + \alpha_i \tag{2.7}$$

for all quantiles τ_j in the interval $(0,1)$. We assume that the individual effect does not represent a distributional shift, since it is unrealistic to estimate it when the number of observations on each individual is small. The individual specific effect α_i is a pure location shift effect on the conditional quantiles of the response.

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

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

where $\rho_{\tau_j}(u) = u(\tau_j - I(u \leq 0))$ is the quantile loss function, and ω_j is a relative weight given to the jth quantile. The weights control the influence of the quantiles on the estimation of the individual effects. For $\lambda = 0$, we have the fixed effects estimator, while for $\lambda > 0$, the penalized estimator with fixed effects. In the analysis, the choice of the shrinkage parameter λ remained to be investigated, therefore we will focus on the derivation and implementation of a method for selecting precisely how much shrinkage it is needed. Under certain conditions, as will become clear in our later analysis, the estimator $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau},\lambda)$ is asymptotically unbiased for all positive λ, therefore it is reasonable to consider choosing λ to minimize asymptotic variance. Our optimal choice of λ will give the minimum variance estimator in the class of penalized quantile regression estimators, the analog of the GLS in the class of penalized least squares estimators for panel data.

3. Asymptotic Theory

We begin with the asymptotics in the case of one quantile when the number of cross sectional units and the number of time periods both go to infinity. Then, we derive the asymptotic distribution of the panel data quantile regression estimators for several quantiles simultaneously estimated. We use notation that is standard. The symbol "$\to$" signifies convergence in probability, "$\rightsquigarrow$" denotes convergence in distribution, sgn(a) is the sign of a, and tr$A$ means trace of the matrix A.

3.1. **Asymptotics for the Unobserved Effects Model when T, and N tend to infinity.** Consider the model,

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it},$$

where y_{it} is the response, $\boldsymbol{x}_{it}$ is the vector of covariates, α_i is a latent time-invariant effect, and u_{it} is the error term. Assume the following regularity conditions,

A 1. The variables y_{it} are independent with conditional (on $\boldsymbol{x}_{it}$, and α_i) distribution functions F_{it}, and continuous densities f_{it} uniformly bounded away from 0 and ∞ at the points $\xi_{it}(\tau)$ for $t = 1, \dots, T$ and $i = 1, \dots, N$.

A 2. The random variables α_i, stochastically independent of $\boldsymbol{x}_{it}$, are exchangeable, identically, and independently distributed with unconditional distribution functions G_i with median zero, and continuous densities g_i for $i = 1, \dots, N$.

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

$$\begin{aligned} \boldsymbol{D}_0 &= \lim_{T,N\to\infty} \frac{\tau(1-\tau)}{TN}(\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X}); \;\; \boldsymbol{D}_1 = \lim_{T,N\to\infty} \frac{1}{TN}(\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{\Phi}\boldsymbol{M}\boldsymbol{X}) \\ \boldsymbol{D}_2 &= \lim_{T,N\to\infty} \frac{1}{4TN}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}); \;\; \boldsymbol{D}_3 = \lim_{T,N\to\infty} \frac{1}{TN}(\tilde{\boldsymbol{X}}'\boldsymbol{\Psi}\tilde{\boldsymbol{X}}) \end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$, $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}$, $\tilde{\boldsymbol{X}} = (\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{X}$, $\boldsymbol{\Phi} = \mathrm{diag}(f_{it}(\xi_{it}(\tau)))$, and $\boldsymbol{\Psi} = \mathrm{diag}(g_i(0))$.

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

A 5. There exists a constant $c > 0$ such that $N^c/T \to 0$.

A 6. The regularization parameter $\lambda_T/\sqrt{T} \to \lambda \geq 0$.

The behavior of the conditional density in a neighborhood of $\xi_{it}(\tau)$ is crucial for the asymptotic behavior of the quantile regression estimator. Condition A1 ensures a well-defined asymptotic behavior of the quantile regression estimator. Condition A2 is an additional condition, to the ones assumed in Koenker (2004), needed to model the unobserved effects model. The unobserved specific effects are randomly generated with $g_i(\cdot)$ being a density function. In condition A3, the existence of the limiting form of the positive definite matrices is used to invoke the Lindeberg-Feller Central Limit Theorem. In particular, while $\boldsymbol{D}_0$ and $\boldsymbol{D}_1$, are used in the piece of objective function that corresponds to the standard quantile regression problem, $\boldsymbol{D}_2$ and $\boldsymbol{D}_3$ are used in the piece of objective function that corresponds to the penalty term. Condition A4 is important both for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function. Condition A5 is needed to make sure that the contribution of the remainder term that comes from the Bahadur representation of the individual effects is asymptotically negligible. Condition A6 is required to achieve a square root-n consistency.

Consider the objective function for a single quantile,

$$V_{TN}(\boldsymbol{\delta}) = \sum_{t=1}^{T}\sum_{i=1}^{N}\{\rho_\tau(y_{it} - \xi_{it}(\tau) - \delta_{0i}/\sqrt{T} - \boldsymbol{x}_{it}'\boldsymbol{\delta}_1(\tau)/\sqrt{TN})$$

$$- \rho_\tau(y_{it} - \xi_{it}(\tau))\} + \lambda_T\{\sum_{i=1}^{N}|\alpha_i + \delta_{0i}/\sqrt{T}| - |\alpha_i|\}$$

where $\xi_{it}(\tau) = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau)$. Note that the minimizer of $V_{TN}(\boldsymbol{\delta})$ is

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

Theorem 1. *Under the regularity conditions A1-6, the minimizer of the objective function*

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

where

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

and $\boldsymbol{B}$, and $\boldsymbol{C}$ are zero mean Gaussian independent vectors with covariance matrices $\boldsymbol{D}_0$, and $\boldsymbol{D}_2$, respectively.

Remark 1. The Bahadur representation of the unobserved specific effects contains only one "interesting" term, which depends on $\boldsymbol{\delta}_1(\tau)$. Lemmas 1 and 2 show that the asymptotic contribution of the terms that contain $\psi(\cdot)$ are negligible confirming that the finite dimensional convergence holds, without the necessity of cumbersome algebra. For completeness, we extended the proof to include the remainder terms in Appendix A.

Corollary 1. *Let $\boldsymbol{\Sigma}_0(\lambda) = \boldsymbol{D}_0 + \lambda^2\boldsymbol{D}_2$, and $\boldsymbol{\Sigma}_1(\lambda) = \boldsymbol{D}_1 + 2\lambda\boldsymbol{D}_3$. Under the conditions of Theorem 1,*

$$\sqrt{TN}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_1(\lambda)^{-1}\boldsymbol{\Sigma}_0(\lambda)\boldsymbol{\Sigma}_1(\lambda)^{-1})$$

Before turning to the penalized estimator, we consider the fixed effects case setting λ equal to zero. In the iid case $\{u_{it}\} \sim F$, we obtain an estimator that is asymptotically similar to the classical fixed effects estimator,

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

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

Corollary 2. *Under the conditions of Theorem 1, the penalized quantile regression estimator $\hat{\boldsymbol{\beta}}(\tau, \lambda^*)$ is asymptotically normally distributed with mean $\boldsymbol{\beta}(\tau)$ and covariance matrix $\boldsymbol{\Sigma}_1(\lambda^*)^{-1}\boldsymbol{\Sigma}_0(\lambda^*)\boldsymbol{\Sigma}_1(\lambda^*)^{-1}$ where $\lambda^* = \arg\min\{tr\boldsymbol{\Sigma}_1(\lambda)^{-1}\boldsymbol{\Sigma}_0(\lambda)\boldsymbol{\Sigma}_1(\lambda)^{-1}\}$.*

3.2. **Asymptotics for the Unobserved Effects Model when T, and N tend to infinity: J Quantiles Simultaneously Estimated.** Consider the following regularity conditions,

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

B 2. The random variables α_i, stochastically independent of $\boldsymbol{x}_{it}$, are exchangeable, identically, and independently distributed with unconditional distribution function G_i with median zero, and continuous densities g_i for $i = 1, \ldots, N$.

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

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

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

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

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

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

B 4. $\max||\boldsymbol{x}_{it}||/\sqrt{TN} \to 0$.

B 5. There exists a constant $c > 0$ such that $N^c/T \to 0$.

B 6. The regularization parameter $\lambda_T/\sqrt{T} \to \lambda \geq 0$.

Condition B1 ensures a well defined asymptotic behavior of the estimators by imposing variability around the J conditional quantiles. Condition B2 is interpreted as A2. The existence of the limiting form of the positive definite matrices, assumed in B3, is needed to invoke the Lindeberg-Feller Central Limit Theorem. Positive definite matrices $\boldsymbol{H}_0$ and $\boldsymbol{H}_1$ are used in the piece of the objective function that corresponds to multiple quantile regression problem, while $\boldsymbol{H}_2$, and $\boldsymbol{H}_3$ are used in the piece of the objective function that corresponds to the penalty term. Conditions B4-6 have the same implications for the asymptotic behavior of panel data quantile regression estimators than A4-6.

Theorem 2. *Under the regularity conditions B1-6, the minimizer of the objective function*

$$argminV_{TN}(\boldsymbol{\delta}_1) \rightsquigarrow argminV_0(\boldsymbol{\delta}_1)$$

where

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

and $\boldsymbol{B}$*, and* $\boldsymbol{C}$ *are zero mean Gaussian independent vectors with covariance matrices* $\boldsymbol{H}_0$*, and* $\boldsymbol{H}_2$*, respectively.*

Corollary 3. *Let* $\boldsymbol{\Gamma}_0(\lambda) = \boldsymbol{H}_0 + \lambda^2\boldsymbol{H}_2$*, and* $\boldsymbol{\Gamma}_1(\lambda) = \boldsymbol{H}_1 + 2\lambda\boldsymbol{H}_3$*. Under the conditions of Theorem 2,*

$$\sqrt{TN}(\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}) - \boldsymbol{\beta}(\boldsymbol{\tau})) \rightsquigarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1})$$

Remark 2. The contribution of the penalty to the limiting objective function leads to a variation of the existing quantile regression asymptotic theory. When the α's are drawn from a conditional distribution function with median zero, the asymptotic bias of the penalized quantile regression estimator is zero, and the asymptotic variance depends on λ.

Corollary 4. *Under the conditions of Theorem 2, the penalized quantile regression estimator* $\hat{\boldsymbol{\beta}}(\boldsymbol{\tau}, \lambda^*)$ *is asymptotically normally distributed with mean* $\boldsymbol{\beta}(\boldsymbol{\tau})$ *and covariance matrix* $\boldsymbol{\Gamma}_1(\lambda^*)^{-1}\boldsymbol{\Gamma}_0(\lambda^*)\boldsymbol{\Gamma}_1(\lambda^*)^{-1}$ *where* $\lambda^* = \arg\min\{tr\boldsymbol{\Gamma}_1(\lambda)^{-1}\boldsymbol{\Gamma}_0(\lambda)\boldsymbol{\Gamma}_1(\lambda)^{-1}\}$.

Corollaries 2 and 4 state the main theoretical results: the asymptotic distribution of the quantile regression estimators. In the next section, we show the existence and uniqueness of λ^*.

4. Existence, Uniqueness and Estimation

The primary objective is now to show that the optimal tuning parameter λ^* exists and it is unique under the regularity conditions. The result, formally established in Theorem 3, holds under the conditions of Theorem 2 and Lemmas 3-6 (Appendix A)[1].

Theorem 3. *Under conditions B1-6, there exists a unique minimizer*

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

Theorem 3 shows that it is possible to have an estimator that is, on average, more efficient than the fixed effects and pooled quantile regression estimators. Unfortunately, the optimal parameter λ^* may be subject to a scaling problem, which can be eliminated selecting one element of the main diagonal of the asymptotic covariance matrix[2]. The following corollary derives the exact amount of shrinkage that leads to the minimum variance quantile regression estimator for a single parameter, say β_k.

B 7. The kth column of $\boldsymbol{X}$ is a time-varying covariate $\boldsymbol{x}_k = (x_{11,k}, ..., x_{NT,k})'$ such that $x_{it,k} \neq x_{is,k}$ for at least for one $s \neq t$.

[1]Lemmas 3 and 4 show that there exists a non-empty and compact set, defined by the eigenvalues ζ of the asymptotic covariance matrix. Lemmas 5 and 6 show that the rational functions, obtained after we derive the spectral decomposition of the asymptotic covariance matrix, are strictly convex.

[2]This was pointed out by Professor B. Hansen.

Corollary 5. *Under conditions B1-7, the unique optimal value of the regularization parameter that minimizes the asymptotic variance of the penalized quantile regression estimator* $\hat{\beta}_k(\tau_j, \lambda^*)$ *is,*

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

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

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

4.1. **Estimation.** This section briefly suggests how to construct an estimator, $\hat{\lambda}$, for the optimal degree of shrinkage, λ^*, that crucially depends on estimation of the conditional density $f(\xi(\tau))$ and the density of the individual effects $g(0)$.

4.1.1. *Estimating* f. The estimates are based upon the large literature on estimating $f(F(\tau)^{-1})$ in the quantile regression model for cross-sectional data (see, e.g., section 3.4 in Koenker (2005) for a review of methods). In iid settings, we need to estimate the (nuisance) sparsity parameter

$$s(\tau) = [f(F(\tau)^{-1})]^{-1}$$

to estimate the precision of the quantile regression estimator. Writing $F(F(t)^{-1}) = t$ and differentiating gives the sparsity $s(t)$, which have been suggested as approaches to estimation, including finding the slope of an empirical based F at the τ-th quantile (Koenker and Bassett 1982). In particular, we use the fixed effects residuals

$$\hat{u}_{it}(\tau) = y_{it} - \boldsymbol{x}_{it}' \hat{\boldsymbol{\beta}}(\tau, 0) - \hat{\alpha}_i(0) \tag{4.1}$$

to construct a first empirical distribution function F, and then an estimate of $s(\tau)$. In non-iid settings, we consider Hendricks and Koenker (1992) proposal estimating the conditional density f at $\boldsymbol{x}$ by

$$\hat{f}_i(\xi(\tau)|\boldsymbol{x}) = 2h_{NT}/\boldsymbol{x}'(\hat{\boldsymbol{\beta}}(\tau + h_{NT}, 0) - \hat{\boldsymbol{\beta}}(\tau - h_{NT}, 0))$$

where h_{NT} is a bandwidth that tends to zero as the sample size increases. We will also consider Powell's (1991) kernel method on normalized residuals, as described in equation (4.1), using the bandwidth and the estimates of scale proposed in Koenker (2005).

4.1.2. *Estimating* g. Here we propose to obtain the estimate $\hat{g}$ based on individual effects $\hat{\alpha}_i(0)$, denoted below simply as $\hat{\alpha}_i$. If this idea could give us a consistent estimate of g, the first asymptotic requirement should be that $\hat{\alpha}_i \to \alpha_i$. The Bahadur representation of this individual "fixed" effect, considered in Koenker (2004), can be written as,

$$\hat{\alpha}_i - \alpha_i = \boldsymbol{A}_2^{-1} \boldsymbol{A}_1 \boldsymbol{A}_0^{-1} \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{x}_{it} \psi_\tau(y_{it} - \xi_{it}(\tau)) + \boldsymbol{A}_2^{-1} \frac{1}{T} \sum_{t=1}^{T} \psi_\tau(y_{it} - \xi_{it}(\tau)) + \boldsymbol{A}_2^{-1} R_{T_i}$$

where $\boldsymbol{A}_0 = (NT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} f_{it}\boldsymbol{x}_{it}\boldsymbol{x}'_{it}$, $\boldsymbol{A}_1 = T^{-1}\sum_{t=1}^{T} f_{it}\boldsymbol{x}'_{it}$, and

$$\boldsymbol{A}_2 \quad = \quad \frac{1}{T}\sum_{t=1}^{T} f_{it} + \boldsymbol{A}_1\boldsymbol{A}_0^{-1}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} f_{it}\boldsymbol{x}_{it}$$

Under the conditions of Koenker's (2004) Theorem 1, and previous regularity conditions, it can be shown that the fixed effects estimator $\hat{\alpha}_i \to \alpha_i$. (This should be regarded as a pure heuristic argument, since we overlook the contribution of the remainder terms). The second minimal requirement should be $\hat{g} \to g$, so below we consider consistent estimation strategies for the density $g(\hat{\alpha}_i)$.

The first approach examines logspline density estimation (Kooperberg and Stone 1991) that employs the exponential family model,

$$g(\alpha;\boldsymbol{\theta}) = \exp\left\{\boldsymbol{\theta}' B(\alpha) - \log\left(\int \exp(\boldsymbol{\theta}' B(\alpha))d\alpha\right)\right\}$$

where $B = (B_1(\alpha), ..., B_p(\alpha))$ are a $\mathcal{B}$-spline basis of S, a p-dimensional subspace of the set of twice continuously differentiable functions on $\mathbb{R}$, and $\boldsymbol{\theta} = (\theta_1, ..., \theta_p) \in \mathbb{R}^p$ with θ_1 and θ_p less than zero. They refer to the family $g(\cdot;\boldsymbol{\theta})$ as a log-spline family, then

$$\hat{g}(\hat{\alpha}_i;\hat{\boldsymbol{\theta}})$$

is the log-spline density estimate. The estimates $\hat{\alpha}_i$ and $\hat{\boldsymbol{\theta}}$ are obtained using fixed effects quantile regression and maximum likelihood (ML), respectively. In logspline models, the main problem is selecting the number of knots, which is analogous to select a bandwidth in kernel density estimation. The density g is estimated using the Kooperberg and Stone knot addition and deletion algorithm.

We also estimate the density of the individual effects using kernel methods,

$$\frac{1}{Nh_N}\sum_{i=1}^{N} K\left(\frac{\hat{\alpha}_i}{h_N}\right)$$

with a sample of individual fixed effects estimates $\{\hat{\alpha}_1, \ldots, \hat{\alpha}_N\}$, a Gaussian kernel $K(\cdot)$ and a bandwidth h_N. The selection of the smoothing parameter is unclear, so we consider (i) Silverman's (1986) rule of thumb $h_N = 0.9\min\{\hat{\sigma}, \mathrm{R}/1.34\}N^{-1/5}$, where $\hat{\sigma}$ is the standard deviation and R denotes the inter-quartile range, (ii) Scott's (1992) variation of the previous formula replacing the constant 0.9 by 1.06, and (iii) Scott's (1992) unbiased cross validation criterion as described in Venables and Ripley (2002).

5. Monte Carlo Results

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

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

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it} \tag{5.1}$$

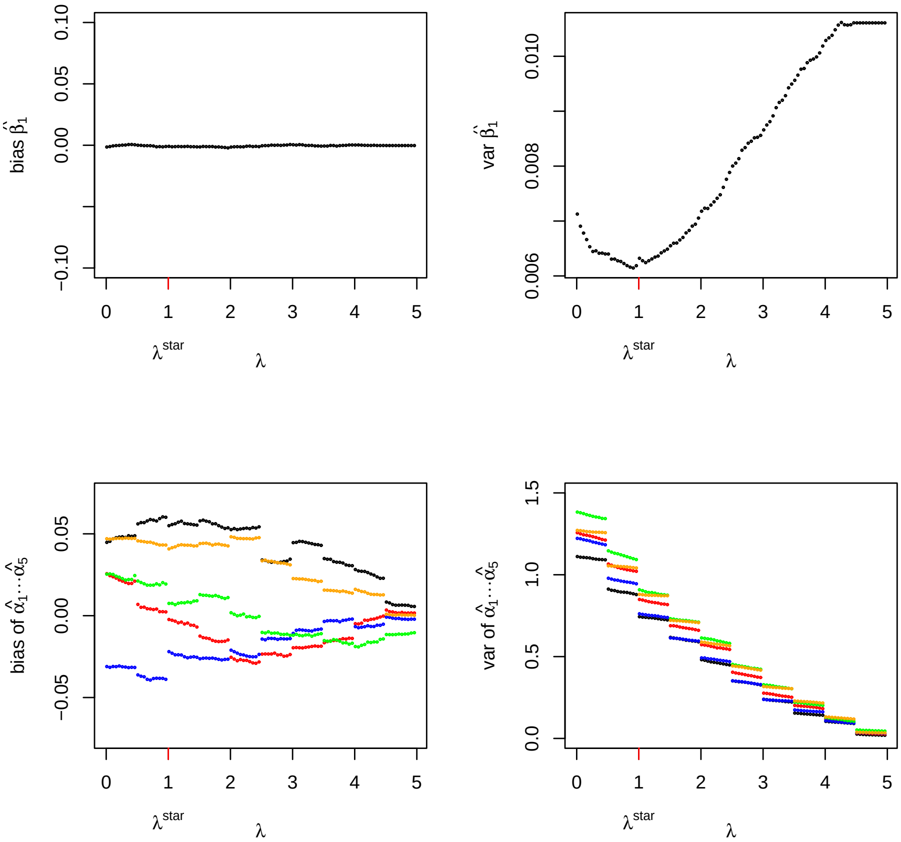

FIGURE 5.1. *Small sample performance of the quantile regression estimators as a function of the regularization parameter. The upper panels present the performance of the slope estimate, and the lower panels the performance of an estimate of the first five individual specific effects. Each dot represents a statistic based on 400 randomly generated samples. The additional vertical line on the x-axis gives the optimal value* λ^*.

and from the location-scale shift model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \gamma_0(1 + \gamma_1 x_{it})u_{it} \tag{5.2}$$

where $\alpha_i \sim G$ and $u_{it} \sim F$. The independent variable x_{it} is generated as in Koenker (2004) and the parameters of the model are assumed to be 0 and 1 for the intercept and slope, respectively.

Considering the location-shift model and assuming that α_i and u_{it} are iid $\mathcal{N}(0,1)$, we evaluate the small sample performance of λ^* generating data from a small data set: $N = 50$ and $T = 5$. Because α_i and u_{it} are independent and identically distributed Gaussian random variables with scale parameter equal to 1, a priori we expect that a tuning parameter equal to 1 would perform well.

Monte Carlo evidence, presented in Figure 5.1, suggests that the estimator of the slope parameter has essentially zero bias for all λ, and its variance does change with λ. The variability of the estimator decreases first, then increases, and lastly it is constant in λ giving rise to the questions: What is the optimal λ parameter, and how can it be determined in this case? Letting $j = k = 2$ and $\tau_2 = \omega_2 = 0.5$, Corollary 5 gives $\lambda^* = 1$. Therefore, the method correctly chooses the value of the shrinkage parameter, reducing the variability of the fixed effect estimator 16 percent without sacrificing bias.

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

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

We estimate λ^* assuming that the distribution function of the error term F and the distribution function of the individual specific effects G are Gaussian and unknown. Without loss of generality, we assume that the parameters of the model are equal to zero. While we estimate the scalar sparsity parameter in the location shift model, we use Powell's kernel method (implemented in Koenker (2006)) in the location-scale shift model. On the other hand, we use a logspline method and a Gaussian kernel with three different bandwidths to estimate g: Silverman's (1986) rule of thumb (S86 in Table 5.1), Scott's (1992) variation (S92), and Scott's (1992) unbiased cross validation (UCV).

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

We compare, in Table 5.3, the performance outside the class of quantile regression estimators considering the following estimators: (1) the ordinary least squares (OLS); (2) the fixed effects estimator (FE);(3) the GLS estimator; (4) the pooled quantile regression estimator (QR); (5) the fixed effect quantile regression estimator (FEQR); (6) the penalized quantile regression (PQR).

We expand the design of the experiment. The y variables are generated as before, and the x series are generated using a similar method to that of Nerlove (1971) and Baltagi

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

TABLE 5.1. *Bias and RMSE of $\hat{\lambda}$ for the optimal amount of shrinkage λ^*.*

(1981), having an intraclass correlation coefficient $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_u^2)$ either $\{0.4, 0.8\}$ for a fixed number of cross sectional and time data points $N = 50$ and $T = 5$, a small panel data

		Bias				RMSE			
		QR	QRFE	PQR	FPQR	QR	QRFE	PQR	FPQR
N	T	Quantile $\tau = 0.25$ - Location Shift Model							
50	5	0.0015	0.0062	0.0049	0.0048	0.0765	0.0793	0.0649	0.0665
50	15	0.0005	0.0005	0.0006	0.0009	0.0636	0.0434	0.0412	0.0410
50	25	0.0005	0.0016	0.0013	0.0012	0.0604	0.0323	0.0304	0.0303
100	5	0.0037	0.0056	0.0044	0.0039	0.0495	0.0575	0.0435	0.0441
100	15	0.0035	0.0014	0.0010	0.0012	0.0451	0.0307	0.0281	0.0278
100	25	0.0009	0.0001	0.0004	0.0005	0.0383	0.0231	0.0216	0.0215
500	5	0.0002	0.0005	0.0005	0.0005	0.0229	0.0276	0.0206	0.0212
500	15	0.0002	0.0002	0.0003	0.0003	0.0190	0.0140	0.0126	0.0126
500	25	0.0004	0.0002	0.0002	0.0002	0.0183	0.0110	0.0103	0.0103
N	T	Quantile $\tau = 0.50$ - Location Shift Model							
50	5	0.0011	0.0025	0.0008	0.0001	0.0709	0.0745	0.0631	0.0626
50	15	0.0004	0.0015	0.0015	0.0017	0.0594	0.0420	0.0385	0.0392
50	25	0.0005	0.0022	0.0017	0.0017	0.0561	0.0306	0.0290	0.0289
100	5	0.0037	0.0068	0.0063	0.0058	0.0443	0.0552	0.0421	0.0415
100	15	0.0037	0.0029	0.0028	0.0028	0.0414	0.0308	0.0276	0.0275
100	25	0.0019	0.0003	0.0001	0.0000	0.0364	0.0228	0.0210	0.0210
500	5	0.0001	0.0008	0.0002	0.0001	0.0206	0.0268	0.0190	0.0189
500	15	0.0002	0.0005	0.0005	0.0005	0.0180	0.0136	0.0125	0.0125
500	25	0.0008	0.0000	0.0000	0.0000	0.0173	0.0109	0.0100	0.0100
N	T	Quantile $\tau = 0.25$ - Location-Scale Shift Model							
50	5	0.0208	0.0173	0.0136	0.0140	0.0796	0.0812	0.0663	0.0656
50	15	0.0172	0.0005	0.0014	0.0014	0.0653	0.0427	0.0385	0.0383
50	25	0.0232	0.0045	0.0050	0.0049	0.0605	0.0289	0.0275	0.0275
100	5	0.0256	0.0127	0.0121	0.0117	0.0511	0.0513	0.0406	0.0411
100	15	0.0213	0.0073	0.0061	0.0059	0.0444	0.0273	0.0254	0.0254
100	25	0.0225	0.0027	0.0026	0.0025	0.0448	0.0229	0.0216	0.0217
500	5	0.0245	0.0144	0.0123	0.0122	0.0235	0.0251	0.0202	0.0201
500	15	0.0229	0.0045	0.0044	0.0044	0.0191	0.0131	0.0119	0.0119
500	25	0.0194	0.0028	0.0026	0.0026	0.0190	0.0090	0.0085	0.0085
N	T	Quantile $\tau = 0.50$ - Location-Scale Shift Model							
50	5	0.0020	0.0041	0.0022	0.0022	0.0721	0.0794	0.0587	0.0597
50	15	0.0020	0.0033	0.0017	0.0016	0.0617	0.0422	0.0378	0.0382
50	25	0.0032	0.0018	0.0020	0.0023	0.0572	0.0288	0.0276	0.0278
100	5	0.0015	0.0032	0.0003	0.0005	0.0481	0.0501	0.0410	0.0407
100	15	0.0002	0.0024	0.0007	0.0006	0.0400	0.0257	0.0253	0.0259
100	25	0.0001	0.0000	0.0001	0.0001	0.0433	0.0225	0.0210	0.0210
500	5	0.0012	0.0003	0.0010	0.0010	.0227	0.0245	0.0192	0.0190
500	15	0.0012	0.0001	0.0002	0.0002	0.0183	0.0126	0.0117	0.0116
500	25	0.0018	0.0004	0.0001	0.0001	0.0181	0.0089	0.0087	0.0087

TABLE 5.2. *Bias and Root Mean Square Error for the Slope. PQR stands for the unfeasible penalized estimator $\hat{\beta}_1(\tau, \lambda^*)$ and FPQR denotes the Feasible Penalized Estimator $\hat{\beta}_1(\tau, \hat{\lambda})$.*

ρ	λ^*	α	u	Statistic	Panel Data Estimators					
					Least Squares			Quantile Regression		
					OLS	FE	GLS	QR	FEQR	PQR
					Quantile $\tau = 0.5$ - Location Shift Model					
0.4	1.45	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0016	0.0000	0.0010	0.0008	0.0015	0.0017
				Std. Error	0.0397	0.0548	0.0361	0.0442	0.0596	0.0433
	0.70	t_3	t_3	Bias	0.0011	0.0018	0.0013	0.0030	0.0018	0.0010
				Std. Error	0.0903	0.0923	0.0744	0.0530	0.0700	0.0495
	0.95	t_3	$\mathcal{N}$	Bias	0.0011	0.0020	0.0014	0.0030	0.0017	0.0034
				Std. Error	0.0674	0.0983	0.0596	0.0720	0.1112	0.0676
	0.70	$\mathcal{N}$	t_3	Bias	0.0074	0.0039	0.0062	0.0055	0.0019	0.0026
				Std. Error	0.0616	0.0782	0.0519	0.0651	0.0633	0.0504
0.8	0.45	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0001	0.0017	0.0012	0.0016	0.0020	0.0025
				Std. Error	0.0473	0.0322	0.0285	0.0547	0.0357	0.0331
	0.20	t_3	t_3	Bias	0.0034	0.0058	0.0052	0.0031	0.0010	0.0015
				Std. Error	0.0884	0.0539	0.0485	0.0668	0.0426	0.0390
	0.45	t_3	$\mathcal{N}$	Bias	0.0022	0.0005	0.0009	0.0027	0.0006	0.0004
				Std. Error	0.0678	0.0423	0.0378	0.0566	0.0483	0.0409
	0.45	$\mathcal{N}$	t_3	Bias	0.0026	0.0010	0.0001	0.0079	0.0047	0.0050
				Std. Error	0.1290	0.0843	0.0740	0.1523	0.0668	0.0630
					Quantile $\tau = 0.5$ - Location-Scale Shift Model					
0.4	1.85	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0010	0.0002	0.0009	0.0010	0.0029	0.0009
				Std. Error	0.0415	0.0607	0.0388	0.0479	0.0650	0.0467
	0.35	t_3	t_3	Bias	0.0029	0.0035	0.0032	0.0065	0.0029	0.0018
				Std. Error	0.1180	0.0742	0.0735	0.0556	0.0549	0.0440
	2.10	t_3	$\mathcal{N}$	Bias	0.0061	0.0029	0.0056	0.0053	0.0063	0.0027
				Std. Error	0.0841	0.1222	0.0745	0.0775	0.1266	0.0727
	1.35	$\mathcal{N}$	t_3	Bias	0.0058	0.0051	0.0048	0.0119	0.0021	0.0057
				Std. Error	0.0994	0.1532	0.0886	0.1030	0.1257	0.0891
0.8	0.25	$\mathcal{N}$	$\mathcal{N}$	Bias	0.0038	0.0029	0.0007	0.0033	0.0042	0.0039
				Std. Error	0.0485	0.0380	0.0297	0.0551	0.0403	0.0357
	0.05	t_3	t_3	Bias	0.0002	0.0024	0.0024	0.0023	0.0003	0.0004
				Std. Error	0.0656	0.0130	0.0123	0.0477	0.0095	0.0095
	0.85	t_3	$\mathcal{N}$	Bias	0.0021	0.0055	0.0044	0.0006	0.0068	0.0036
				Std. Error	0.0644	0.0443	0.0367	0.0532	0.0459	0.0392
	0.55	$\mathcal{N}$	t_3	Bias	0.0306	0.0066	0.0011	0.0457	0.0040	0.0085
				Std. Error	0.2502	0.1472	0.1186	0.2707	0.1154	0.1080

TABLE 5.3. *Small Sample Performance of Panel Data Slope Estimators.*

set representing an unfavorable scenario for λ selection. The parameters of the models are assumed to be 0 and 1 for the intercept and the slope, respectively[3]. We report the results based on 400 replications.

Motivated by Horowitz and Markatou (1996) finding of student t tail behavior, we investigate the relative performance outside the Gaussian setting considering three alternatives to the standard case of error components normally distributed ($\mathcal{N}$). The variables are

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

drawn from a t-student distribution with 3 degrees of freedom (t_3), and also from different distributions (e.g., α_i is drawn from $\mathcal{N}$ and u_{it} from t_3)[4].

In Table 5.3, we see that the PQR is unbiased, robust, and it attains the minimum variance in the class of quantile regression estimators for longitudinal data. The maximum bias is 0.5 percent, and the estimator is more efficient than the true GLS estimator when $u_{it} \sim t_3$. When the intraclass correlation coefficient is positive, the penalized quantile regression estimator is considerably more efficient than both QR and FEQR.

6. Two Simple Examples

In this section, we use data from Baltagi (2001) and Cameron and Trivedi (2005) to investigate the performance of the λ selection device in economic applications. The first example reproduces Baltagi's study of the relationship between public capital and economic performance, the second application complements Cameron and Trivedi's estimates of the intertemporal substitution elasticity of labor-supply (MaCurdy (1981), Altonji (1986), Ziliak (1997), among others). Our objective is to demonstrate how the penalized quantile regression estimator for panel data can be obtained and employed.

6.1. **Public Capital Productivity.** We briefly reconsider an example in Baltagi's (2001) book to compare estimates and illustrate how the estimated asymptotic variance could be obtained. The model is,

$$y_{it} = \beta_0 + \beta_1 k_{1,it} + \beta_2 k_{2,it} + \beta_3 l_{it} + \beta_4 U_{it} + \alpha_i + u_{it}$$

where y_{it} is the log of gross state product, $k_{1,it}$ is the log of public capital, $k_{2,it}$ is the log of private capital, l_{it} is the log of labor input, U_{it} is state unemployment, and α_i is a state fixed effect. This panel consists of annual observations for 48 US states over the period 1970-1986.

We first estimated the asymptotic covariance matrix using the methods presented above, and then we determined that $\hat{\lambda}$ equal to 0.47 minimizes an average of the estimated asymptotic variances of the slopes at the median (Figure 6.1, panels a-d). We observe that the penalized estimator is in general more efficient than the within estimator and less efficient than the classical random effects estimator. The proposed λ selection device, which is computationally attractive relative to the resampling methods that are being investigated, could be also considered to estimate the standard error of the fixed effects estimator by simply evaluating the estimated covariance matrix at λ equal to zero.

6.2. **Hours and Wages.** We now reconsider the example presented in Cameron and Trivedi's (2005) that uses Gaussian panel data estimators to estimate the responsiveness of hours worked to hourly wage. Although the fixed effects is in principle the preferred estimator, tests suggest that classical penalized estimators could be used. In the following simple framework, we try to explain why it may be informative to consider the effect of wages in the lower quantiles of the conditional distribution of hours.

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

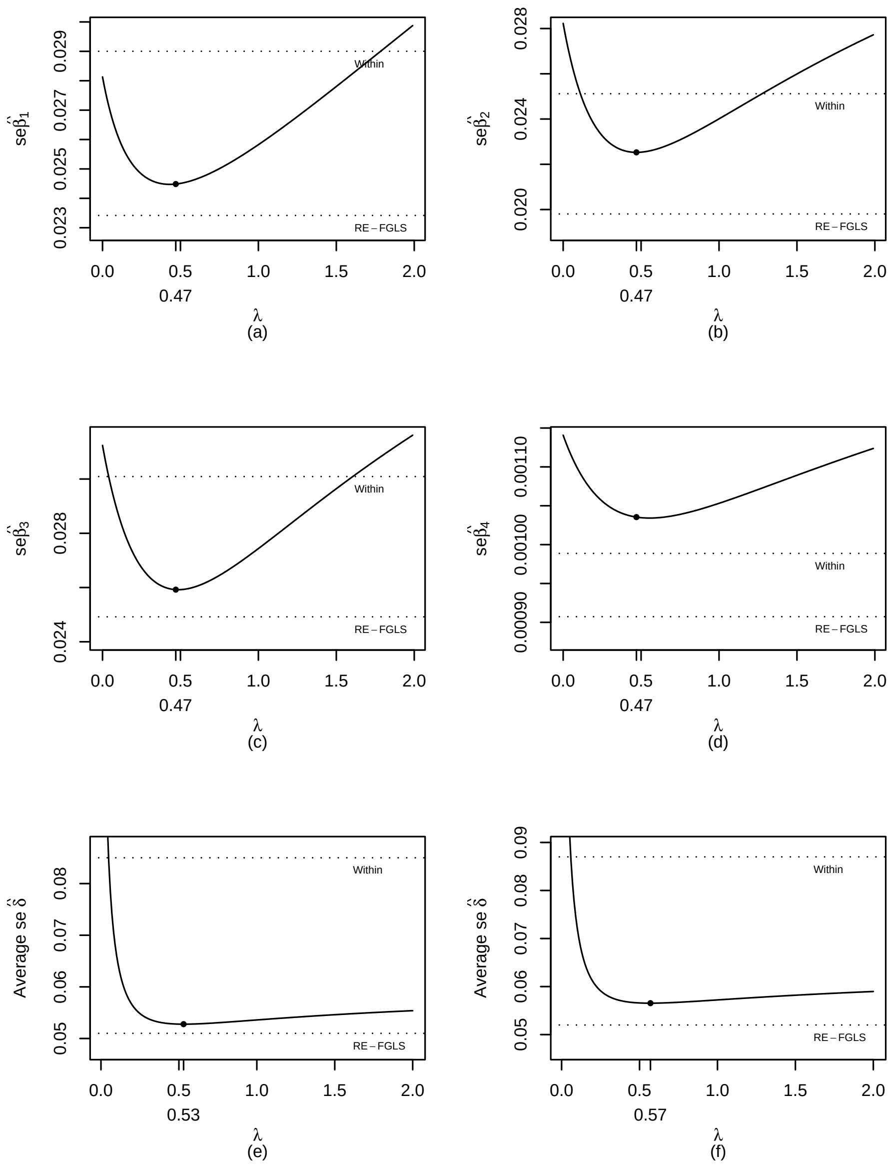

FIGURE 6.1. *Profile of Estimated Asymptotic Variance as the Shrinkage Parameter Changes. The dotted lines give the standard error of the least squares fixed effects and the feasible GLS estimators. The continuous lines are the standard errors of the penalized estimator as a function of λ, and the additional vertical line on the x-axis denotes $\hat{\lambda}$. While panels (a)-(d) corresponds to the example presented in Section 6.1, panels (e)-(f) corresponds to the example presented in Section 6.2.*

6.2.1. *A framework for additional empirical research.* We consider the classical life-cycle model of consumption and labor supply (MaCurdy (1981), Altonji (1986), Ziliak (1997)), assuming the following convenient additively separable utility function

$$U(c_t, h_t) = c_t^{\nu_1} - s_t h_t^{\nu_2}$$

where $0 < \nu_1 < 1$, $\nu_2 > 1$, c is consumption, h is hours of work at age t, and s is a taste shifter that may depend on socioeconomic variables including age. 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

$$\ln(h_{it}) = \alpha_{i0} + \delta \ln(w_{it}) + \gamma t - \delta \ln(s_{it}) \tag{6.1}$$

where ln denotes natural logarithm, $\delta = (\nu_2 - 1)^{-1}$ is the intertemporal substitution elasticity, and α_{i0} denotes the marginal utility of wealth at time 0 that is correlated with the independent variables.

MaCurdy argued that the "taste-shifter" variables may be included in a regression analysis. He suggested introducing them as a linear term in the labor supply equation, opening the door for the consideration of alternatives. We may want to incorporate, for example, a change in the mean and variance of s over time (e.g., the disutility of work is a function of consumer's health[5]). We introduce the variables by simply considering

$$s_{it} = \exp\{\sigma_i - (1 - \boldsymbol{\eta}' \boldsymbol{x}_{it}) u_{it}\}$$

where $\boldsymbol{x}_{it}$ is a vector of variables that could affect preferences, σ_i is an idiosyncratic component and u_{it} is a zero mean variable. Replacing the equation in (6.1), we obtain,

$$\ln(h_{it}) = \alpha_i + \delta \ln(w_{it}) + \gamma t + \delta(1 - \boldsymbol{\eta}' \boldsymbol{x}_{it}) u_{it} \tag{6.2}$$

If $\boldsymbol{x}_{it}$ includes natural logarithm of hourly wages[6], we obtained a labor-supply function with multiplicative heterocedasticity[7]. A change in wages increases the number of hours supplied by a typical worker, but if his/her actual consumption of leisure is high, it may be the case that this consumer would be able to offer more hours than the average worker.

6.2.2. *Data.* We use Ziliak's (1997) PSID sample, also used in Cameron and Trivedi (2005), which includes 5320 observations over ten years: 1978-1987. The sample, which is similar to other data used in previous labor supply studies, includes 532 married, working men aged between 22 and 60 in 1978. The data set includes observations on annual hours worked (the mean is 2182 hours with a standard deviation of 492), hourly wage reported by the

[5]In this PSID sample, while the empirical probability of having health problems is 5 percent with a standard deviation of 0.22 for workers under age 45, it is 9.7 percent with a standard deviation of 0.30 for workers over age 45.

[6]Tipically $\boldsymbol{x}_{it}$ includes number of children and age (e.g. today's disutility of working is nondecreasing in age). We may also argue that today's disutility of working is increasing in number of hours worked in previous weeks, which is correlated with hourly wage. For simplicity we assume that $\boldsymbol{\eta}$ is a vector of ones.

[7]We test for heterocedasticity in the independent variables using the models that we will estimate below. We regress natural logarithm of annual hours worked on natural logarithm of wage, age, number of kids, and a dummy variable for bad health. Like Cameron and Trivedi (2005), we reject (at 1 percent level of significance) the hypothesis of spherical error in this data set.

survey participants ($14.9 with a standard deviation of 7.3), age, number of children, and an indicator for bad health.

	Weights	$\hat{\lambda}$	Quantiles 0.05	0.10	0.15	0.25	0.5	Mean
Fixed Effects Methods			Dependent Variable = Log of Annual Hours Worked[a]					
Log of Hourly Wage	No	0	0.123	0.049	0.029	0.012	0.000	0.168
Standard Error			(0.038)	(0.033)	(0.025)	(0.018)	(0.024)	(0.085)
Boot Standard Error			[0.032]	[0.022]	[0.019]	[0.016]	[0.016]	[0.084]
Log of Hourly Wage	Yes	0	0.133	0.056	0.029	0.009	0.000	0.109
Standard Error			(0.197)	(0.282)	(0.160)	(0.079)	(0.041)	(0.084)
Boot Standard Error			[0.032]	[0.022]	[0.019]	[0.016]	[0.016]	[0.083]
Fixed Effects Methods			Dependent Variable = Log of Annual Hours Worked[b]					
Log of Hourly Wage	No	0	0.120	0.052	0.032	0.014	0.004	0.167
Standard Error			(0.035)	(0.032)	(0.027)	(0.022)	(0.022)	(0.087)
Boot Standard Error			[0.035]	[0.026]	[0.025]	[0.023]	[0.023]	[0.085]
Log of Hourly Wage	Yes	0	0.121	0.047	0.017	-0.002	-0.010	0.110
Standard Error			(0.205)	(0.269)	(0.202)	(0.089)	(0.042)	(0.084)
Boot Standard Error			[0.034]	[0.024]	[0.022]	[0.018]	[0.018]	[0.082]
Penalized Methods			Dependent Variable = Log of Annual Hours Worked[a]					
Log of Hourly Wage	No	0.54	0.126	0.049	0.029	0.012	0.000	0.119
Standard Error			(0.031)	(0.029)	(0.023)	(0.017)	(0.022)	(0.051)
Boot Standard Error			[0.032]	[0.023]	[0.021]	[0.019]	[0.019]	[0.056]
Log of Hourly Wage	Yes	0.53	0.128	0.055	0.029	0.009	0.000	0.120
Standard Error			(0.060)	(0.060)	(0.057)	(0.049)	(0.034)	(0.052)
Boot Standard Error			[0.030]	[0.019]	[0.015]	[0.013]	[0.012]	[0.058]
Penalized Methods			Dependent Variable = Log of Annual Hours Worked[b]					
Log of Hourly Wage	No	0.44	0.125	0.054	0.034	0.016	0.007	0.118
Standard Error			(0.031)	(0.029)	(0.025)	(0.021)	(0.021)	(0.052)
Boot Standard Error			[0.034]	[0.025]	[0.023]	[0.021]	[0.021]	[0.056]
Log of Hourly Wage	Yes	0.57	0.131	0.059	0.031	0.010	0.004	0.120
Standard Error			(0.065)	(0.063)	(0.062)	(0.052)	(0.035)	(0.052)
Boot Standard Error			[0.030]	[0.020]	[0.016]	[0.013]	[0.013]	[0.058]

Table 6.1. *Simple Estimates of the Labor Supply Elasticity. Mean refers to within and first difference estimators (Fixed Effects Methods) and GLS and MLE random effects estimators (Penalized Methods). Letter (a) indicates models without covariates and (b) models with age, number of children, and a dummy variable for bad health included as covariates. The total number of observations is 5320.*

6.2.3. *Empirical Results.* Table 6.1 presents results from both the conditional mean model of equation (6.2) (within and first difference estimators in the category "fixed effects methods" and GLS and MLE random effects in the category "penalized methods") and the conditional quantile model for five quantiles $\{0.05, 0.10, 0.15, 0.25, 0.50\}$. We consider unweighted estimators and weighted estimators with weights $\{0.05, 0.10, 0.15, 0.25, 0.5\}$ over the estimated quantiles.

Fixed effects estimates are unbiased if this effect is time invariant and wages are uncorrelated with the error term. The within estimates suggests a positive elasticity of substitution,

	Estimates of λ^*	Quantiles 0.05	0.10	0.15	0.25	0.5
	Dependent Variable = Log of Annual Hours Worked^a					
Standard Error $\hat{\delta}(\tau_j, \tilde{\lambda})$	1.440	0.035	0.032	0.026	0.020	0.023
Standard Error $\hat{\delta}(\tau_j, \hat{\lambda})$	0.540	0.031	0.029	0.023	0.017	0.022
Variance Percent Change		-0.226	-0.172	-0.219	-0.228	-0.101
	Dependent Variable = Log of Annual Hours Worked^b					
Standard Error $\hat{\delta}(\tau_j, \tilde{\lambda})$	1.450	0.038	0.033	0.029	0.023	0.023
Standard Error $\hat{\delta}(\tau_j, \hat{\lambda})$	0.440	0.031	0.029	0.025	0.021	0.021
Variance Percent Change		-0.333	-0.237	-0.239	-0.226	-0.163

TABLE 6.2. *Change in the precision of the penalized estimates of the elasticity parameter. The estimator $\tilde{\lambda}$ is obtained using MLE; (a) stands for omitted and (b) included covariates (age, number of children, and a dummy variable for bad health). The total number of observations is 5320.*

with a point estimate 0.168 and a standard error, corrected for both serial correlation and heteroscedasticity, equal to 0.085. If instead of focusing at the mean effect we consider responsiveness of labor supply to wages at several quantiles, we see a tendency to decrease as we go across the τ's. For example, considering the unweighted estimator (e.g., equal weights over the quantiles), the elasticity changes from 12 percent at the 0.05 quantile to 0 percent at the 0.5 quantile.

Although the theoretical framework suggests that the fixed effects is the preferred estimator, a Hausman test does not reject the null hypothesis that the marginal utility of wealth is independent of log of hourly wages. Therefore, the more efficient penalized estimators *may* be considered in this data set (Cameron and Trivedi (2005))[8].

The penalized estimators crucially depend on parameters estimated in a first stage. For example, in the case of quantile regression, we find $\hat{\lambda}$ by minimizing estimated asymptotic variance of the elasticity estimator $\hat{\delta}$ over the quantiles. In the bottom panels of Figure 6.1, we show two curves representing the average of the estimated variances of the weighted penalized estimator for the elasticity parameter. The smooth curves are obtained using the logspline method and Koenker's (2006) implementation of Powell's (1991) method. While $\hat{\lambda}$ is equal to 0.53 in a model without "taste shifters" (panel e), it is equal to 0.57 in a model

[8]A natural concern may arise about the estimation of λ^* given the possible correlation between the sign of α_i and wages, but the estimator $\hat{\lambda}$ is consistent because it is a function of unbiased estimates of g and f. In other words, even in cases where the individual specific effects and the independent variable may be correlated, we can rely on $\hat{\lambda}$ because it uses fixed effects estimates for the non-parametric estimation of the densities. Of course, we are explicitly neglecting that $\ln(w_{it})$ may be correlated with u_{it} (e.g., a measurement error problem), but this issue is out of the scope of the paper. On the other hand, the point estimates arising from the second stage are affected when the sign of α and wages are not independent. The reason is that the λ selection device ignores the contribution of the bias in the estimation of $\delta(\tau_j)$. Two direction are being explored. The first one incorporates the bias contribution in Theorem 3 to find the new optimal (now, in terms of MSE) degree of shrinkage. The second alternative is to follow Chamberlain (1982) approach decomposing the individual specific effects in two terms. Penalizing the term that is a "pure" noise offers the possibility of reducing the variability of the fixed effects estimator without sacrificing bias.

with age, number of children and a dummy variable for bad health introduced as covariates (panel f).

In the case of classical random effects estimators, we use feasible GLS and maximum likelihood (MLE) to estimate the variances of α_i and u_{it} obtaining the ratio $\hat{\sigma}_u^2/\hat{\sigma}_\alpha^2$. A tentative, more simple approach for λ selection in quantile regression may consider instead $\tilde{\lambda} = \hat{\sigma}_u/\hat{\sigma}_\alpha$, but it could incorrectly estimate the optimal degree of shrinkage because the estimator is not robust to departures from spherical errors. In our PSID sample, this alternative leads to less precise point estimates, and consequently incorrect inference (Table 6.2). Averaging over the estimated quantiles, $\hat{\lambda}$ offers an estimated variance reduction of 19 percent in a model without taste shifters, and it offers an estimated variance reduction of 24 percent in a model that includes age, number of kids and an indicator for bad health as covariates.

In a second stage, we use the values of $\hat{\lambda}$ to obtain point estimates and covariance matrices (Table 6.1). For completeness, we also report estimates of the standard errors using panel-boostrap replacing pairs $\{(\ln(\boldsymbol{h}_i), \ln(\boldsymbol{w}_i)) : i = 1, \ldots, N\}$ over i, as suggested in the literature (Davison and Hinkley (1997)). Therefore, the first standard error in Table 6.1 is computed using $\hat{\lambda}$ and the second one is computed by panel-bootstrap with 1000 replications.

We see that penalized estimates for the elasticity parameter are quite similar to the fixed effects estimates, but there are large differences in terms of precision. For example, we observe that the standard error of the unweighted estimator is reduced 20 percent at the 0.05 quantile and 9 percent at the 0.5 quantile in a model without covariates. The standard error reduction is even more dramatic for weighted estimators.[9]

Quantile regression seems to provide additional information than the conditional mean approach, suggesting that the location-shift model is inappropriate for this data set. Moreover, we obtain more precise estimates of the responsiveness of wages, not only at the center of the distribution, but also in the lower tail. The estimated effects in the lower tail of the conditional distribution of log of annual hours are positive and significant, with a tendency to decrease as we move to the center of the distribution. The evidence suggests that wages and hours are positively correlated, but workers consuming long hours of leisure, given their actual wage rate, may prefer to offer more hours.

7. Conclusions and Extensions

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

[9]Monte carlo evidence and the empirical evidence suggests that the implementation of the λ selection device could give the correct size for the variance of unweighted estimators in panels with 5000 observations, but the evidence also suggests that further research is needed on selecting the weights ω_j.

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

Appendix A. Proofs

The proofs refer to Knight's (1998) identities

$$\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau + \int_0^v (I(v \le s) - I(v \le 0))ds \tag{A.1}$$

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

where $\psi_\tau(u) = \tau - I(u \le 0)$ is the quantile influence function.
Proof of Theorem 1: Following the conditions and argument of Ruppert and Carroll (1980), Koenker and Portnoy (1987), and Koenker (2004), for any $(\Delta_{0i}, \Delta_1) > 0$,

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

where

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

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

$$\begin{aligned}
\mathbb{E}v_i &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(F_{it}\left(\xi_{it}(\tau) + \frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - \tau\right) + 2\frac{\lambda_T}{\sqrt{T}}\left(\frac{1}{2} - G_i\left(-\frac{\delta_{0i}}{\sqrt{T}}\right)\right) \\
&= -\frac{1}{\sqrt{T}} \sum_{t=1}^T f_{it}(\xi_{it}(\tau))\left(\frac{\delta_{0i}}{\sqrt{T}} + \boldsymbol{x}_{it}' \frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}}\right) - 2\frac{\lambda_T}{\sqrt{T}} g_i(0)\frac{\delta_{0i}}{\sqrt{T}} + o(1)
\end{aligned}$$

Letting $f_i = T^{-1}\sum_{t=1}^T f_{it}(\xi_{it}(\tau)) + 2T^{-1}\lambda_T g_i(0)$, we find that

$$\frac{\hat{\delta}_{0i}}{\sqrt{T}} = -\tilde{\boldsymbol{x}}_i' \frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}} + \frac{1}{Tf_i}\sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau)) - 2\frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{T}f_i}\psi_{0.5}(\alpha_i) + \frac{R_{Ti}}{\sqrt{T}}$$

where $\tilde{\boldsymbol{x}}_i = \sum_{t=1}^{T} w_{it}\boldsymbol{x}_{it}$, and $w_{it} = f_{it}(\xi_{it}(\tau))/Tf_i$. By A5-A6, Lemma 1 and Koenker's (2004) Theorem 1, the second, third and four terms are $o(1)$, therefore we write the objective function as

$$\begin{aligned} V_{TN}(\boldsymbol{\delta}_1(\tau)) &= \sum_{t=1}^{T}\sum_{i=1}^{N}\{\rho_\tau(y_{it} - \xi_{it}(\tau) - (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau)/\sqrt{TN}) \\ &- \rho_\tau(y_{it} - \xi_{it}(\tau))\} + \lambda_T\{\sum_{i=1}^{N}|\alpha_i - \tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1(\tau)/\sqrt{TN}| - |\alpha_i|\} \end{aligned}$$

Using Knight's (1998) identity, we decompose the objective function in four parts,

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

where,

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

with $v_{it,TN} = (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau)/\sqrt{TN}$. 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. By the Lindeberg-Feller Central Limit Theorem and conditions A3-4,

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

The second term converges in probability to a quadratic term in $\boldsymbol{\delta}_1(\tau)$. Note that

$$\begin{aligned} \mathbb{E}V_{TN}^{(2)}(\boldsymbol{\delta}_1(\tau)) &= \frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sqrt{TN}\int_0^{v_{it}}(F_{it}(\xi_{it}(\tau) + s/\sqrt{TN}) - F_{it}(\xi_{it}(\tau)))ds \\ &= \frac{1}{2TN}\sum_{t=1}^{T}\sum_{i=1}^{N}f_{it}(\xi_{it}(\tau))((\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau))^2 + o(1) \\ &\to \frac{1}{2}\boldsymbol{\delta}_1(\tau)'\boldsymbol{D}_1\boldsymbol{\delta}_1(\tau) \end{aligned}$$

The variance of $V_{TN}^{(2)}(\boldsymbol{\delta}_1(\tau))$ converges to zero by condition A4.

The last two terms of $V_{TN}(\boldsymbol{\delta}_1(\tau))$ represents a decomposition of the stochastic penalty term

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

By the Lindeberg-Feller Central Limit Theorem, the Slutsky theorem and conditions A3-4, the third term is asymptotically Gaussian,

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

The fourth term $V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau))$ is asymptotically quadratic in $\boldsymbol{\delta}_1(\tau)$,

$$\begin{aligned}\mathbb{E}V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau)) &= 2\frac{\lambda_T}{NT}\sum_{i=1}^{N}\int_0^{\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau)}\sqrt{NT}(G_i(s/\sqrt{TN})-G_i(0))ds \\ &= \frac{\lambda_T}{TN}\sum_{i=1}^{N}g_i(0)(\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau))^2+o(1) \\ &\to \lambda\boldsymbol{\delta}_1(\tau)'\boldsymbol{D}_3\boldsymbol{\delta}_1(\tau)\end{aligned}$$

Using regularity condition A4, we obtain

$$\text{Var}(V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau))) \quad \leq \quad 2\frac{\lambda_T}{\sqrt{TN}}\max|\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau)|\sum_{i=1}^{N}\mathbb{E}V_{TN,i}^{(4)}(\boldsymbol{\delta}_1(\tau)) \to 0$$

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

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

■

Proof of Theorem 2: Following the conditions and argument of Ruppert and Carroll (1980), Koenker and Portnoy (1987), and Koenker (2004), for any $(\Delta_{0i},\Delta_1)>0$,

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

where

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

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

$$\begin{aligned}\mathbb{E}v_{ij} &= -\frac{1}{\sqrt{T}}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j\left(F_{it}\left(\xi_{it}(\tau_j)+\frac{\delta_{0i}}{\sqrt{T}}+\boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}\right)-\tau_j\right)+2\frac{\lambda_T}{\sqrt{T}}\left(\frac{1}{2}-G_i\left(-\frac{\delta_{0i}}{\sqrt{T}}\right)\right) \\ &= -\frac{1}{\sqrt{T}}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}(\xi_{it}(\tau_j))\left(\frac{\delta_{0i}}{\sqrt{T}}+\boldsymbol{x}_{it}'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}\right)-2\frac{\lambda_T}{\sqrt{T}}g_i(0)\frac{\delta_{0i}}{\sqrt{T}}+o(1)\end{aligned}$$

Letting $f_{ij}=T^{-1}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j f_{it}(\xi_{it}(\tau_j))+2T^{-1}\lambda_T g_i(0)$, we find that

$$\frac{\hat{\delta}_{0i}}{\sqrt{T}} = -\sum_{j=1}^{J}\tilde{\boldsymbol{x}}_i(\tau_j)'\frac{\boldsymbol{\delta}_1(\tau_j)}{\sqrt{TN}}+\frac{1}{Tf_{ij}}\sum_{j=1}^{J}\sum_{t=1}^{T}\omega_j\psi_{\tau_j}(y_{it}-\xi_{it}(\tau_j))-2\frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{T}f_{ij}}\psi_{0.5}(\alpha_i)+\frac{R_{Tij}}{\sqrt{T}}$$

where $\tilde{\boldsymbol{x}}_i(\tau_j)=\sum_{t=1}^{T}w_{itj}\boldsymbol{x}_{it}$, and $w_{itj}=f_{it}(\xi_{it}(\tau_j))/Tf_{ij}$. By B5-B6, Lemma 1 and Koenker's (2004) Theorem 1, the terms that do not depend on the parameter of main interest $\boldsymbol{\delta}_1(\boldsymbol{\tau})$ are $o(1)$, therefore the objective function can be written as

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

Again we decompose the equation in four terms

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

where,

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

with $v_{itj,TN}=(\boldsymbol{x}_{it}'-\tilde{\boldsymbol{x}}_i(\tau_j)')\boldsymbol{\delta}_1(\tau_j)/\sqrt{TN}$. The first two terms corresponds to the decomposition of the weighted sum of the 'check' functions, and the last two terms corresponds to the decomposition of the penalty term.

The first term is asymptotically Gaussian. By the Lindeberg-Feller Central Limit Theorem, and conditions B3-4,

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

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

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

The variance of $V_{TN}^{(2)}(\boldsymbol{\delta}_1)$ converges to zero by condition B4. The last two terms of $V_{TN}(\boldsymbol{\delta})$ represents a decomposition of the stochastic penalty term,

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

By the Lindeberg-Feller Central Limit Theorem, the Slutsky Theorem, and conditions B3-4, the third term is asymptotically Gaussian,

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

The last term $V_{TN}^{(4)}(\boldsymbol{\delta}_1)$ has a quadratic contribution,

$$\begin{aligned}\mathbb{E}V_{TN}^{(4)}(\boldsymbol{\delta}_1) \quad &= \quad 2\frac{\lambda_T}{TN}\sum_{i=1}^{N}\int_0^{\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1}\sqrt{NT}(G_i(s/\sqrt{TN})-G_i(0))ds \\ &= \quad \frac{\lambda_T}{TN}\sum_{i=1}^{N}g_i(0)(\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1)^2+o(1) \\ &\to \quad \lambda\boldsymbol{\delta}_1'\boldsymbol{H}_3\boldsymbol{\delta}_1\end{aligned}$$

Using regularity condition B4, we obtain

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

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

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

■

Lemma 1. $\bar{\psi}_{it}(\tau_j) = (Tf_{ij})^{-1}\sum_{t=1}^{T}\omega_j\psi_{\tau_j}(y_{it}-\xi_{it}(\tau_j)) \to 0$ *as* $T\to\infty$.

Proof: Trivially $\mathbb{E}\sum_{j=1}^{J}\bar{\psi}_{it}(\tau_j)=0$. Using Chebyshev inequality,

$$\begin{aligned}P\left\{\left|\sum_{j=1}^{J}\bar{\psi}_{it}(\tau_j)-\mathbb{E}\sum_{j=1}^{J}\bar{\psi}_{it}(\tau_j)\right|\geq\epsilon\right\} \quad &\leq \quad \frac{\sum_{j=1}^{J}\omega_j^2\tau_j(1-\tau_j)+\sum_{j\neq h}^{J}\omega_j\omega_h\tau_j(1-\tau_h)}{T(\epsilon f_{ij})^2} \\ &\to \quad 0\end{aligned}$$

It is straightforward to show that the result holds also for $J=1$ (Theorem 1). ■

Lemma 2. $\bar{\psi}_i(0.5) = (\sqrt{T}f_i)^{-1}\psi_{0.5}(\alpha_i) \to 0$ *as* $T\to\infty$.

Proof: Again $\mathbb{E}\bar{\psi}_i(0.5)=0$. Using Chebyshev inequality,

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

since by A4-A5 and Slutsky Theorem,

$$f_i=\frac{1}{T}\sum_{t=1}^{T}f_{it}+\frac{\lambda_T}{\sqrt{T}}\frac{2g_i}{\sqrt{T}}\to\mathbb{E}f_i+\lambda\,0=\mathbb{E}f_i$$

■

Proof of Remark 1: For completeness, we present a proof that validates the approach used in Theorems 1 and 2. The contribution to the limiting stochastic objective function of the terms that does not depend on $\boldsymbol{\delta}_1(\tau)$ are, under our assumptions, asymptotically negligible. The analysis of Koenker (2004) shows that, under the condition that T grows faster than N, the contribution of R_{Ti} is asymptotically negligible. We will evaluate the contribution of the remainder term to the limiting form of the stochastic objective function.

Consider the objective function that concentrates out the Bahadur representation of the individual specific effects,

$$\begin{aligned} V_{TN}(\boldsymbol{\delta}_1(\tau)) &= \sum_{t=1}^{T}\sum_{i=1}^{N}\{\rho_\tau\left(y_{it} - \xi_{it}(\tau) - (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}} - \frac{\varphi_{it}}{\sqrt{T}} + \frac{\lambda_T}{\sqrt{T}}\frac{\varphi_i}{\sqrt{T}}\right) \\ &- \rho_\tau(y_{it} - \xi_{it}(\tau))\} + \lambda_T\left\{\sum_{i=1}^{N}\left|\alpha_i - \tilde{\boldsymbol{x}}'_i\frac{\boldsymbol{\delta}_1(\tau)}{\sqrt{TN}} + \frac{\varphi_{it}}{\sqrt{T}} - \frac{\lambda_T}{\sqrt{T}}\frac{\varphi_i}{\sqrt{T}}\right| - |\alpha_i|\right\} \end{aligned}$$

where $\varphi_{it} = (\sqrt{T}f_i)^{-1}\sum_{t=1}^{T}\psi_\tau(y_{it} - \xi_{it}(\tau))$, and $\varphi_i = (2f_i)^{-1}\psi_{0.5}(\alpha_i)$.

Using Knight's (1998) identity, we decompose the objective function in four terms.

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

where

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

where $v_{it,TN} = (\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau)/\sqrt{TN} + \varphi_{it}/\sqrt{T} - \lambda_T\varphi_i/T$, and $r_{it,TN} = -\tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1(\tau)/\sqrt{TN} + \varphi_{it}/\sqrt{T} - \lambda_T\varphi_i/T$.

The first term is asymptotically Gaussian. Using the Lindeberg-Feller Central Limit Theorem and conditions A3-6, we obtain that

$$\begin{aligned} V_{TN}^{(1)}(\boldsymbol{\delta}_1(\tau)) &= -\sum_{t=1}^{T}\sum_{i=1}^{N}\psi_\tau(y_{it} - \xi_{it}(\tau))\left((\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau)/\sqrt{TN} + \varphi_{it}/\sqrt{T} - \lambda_T\varphi_i/T\right) \\ &\rightsquigarrow -\boldsymbol{\delta}_1(\tau)'\boldsymbol{B} \end{aligned}$$

Under condition A5, by the continuous mapping theorem, the second term is

$$\begin{aligned} \mathbb{E}V_{TN}^{(2)}(\boldsymbol{\delta}_1(\tau)) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sqrt{T}\int_0^{v_{it}}\left(F_{it}\left(\xi_{it}(\tau) + s/\sqrt{T}\right) - F_{it}(\xi_{it}(\tau))\right)ds \\ &= \frac{1}{2T}\sum_{t=1}^{T}\sum_{i=1}^{N}f_{it}(\xi_{it}(\tau))\left((\boldsymbol{x}'_{it} - \tilde{\boldsymbol{x}}'_i)\boldsymbol{\delta}_1(\tau)/\sqrt{N} + \varphi_{it} - \left(\lambda_T/\sqrt{T}\right)\varphi_i\right)^2 + o(1) \\ &\to \frac{1}{2}\boldsymbol{\delta}_1(\tau)'\boldsymbol{D}_1\boldsymbol{\delta}_1(\tau) \end{aligned}$$

The third term, by the Lindeberg Central Limit Theorem, Slutsky Theorem and conditions A3-6, is asymptotically equal to,

$$V_{TN}^{(3)}(\boldsymbol{\delta}_1(\tau)) = -\lambda_T\sum_{i=1}^{N}\text{sgn}(\alpha_i)\left(\tilde{\boldsymbol{x}}'_i\boldsymbol{\delta}_1(\tau)/\sqrt{NT} + \varphi_{it}/\sqrt{T} - \lambda_T\varphi_i/T\right) \rightsquigarrow -\lambda\boldsymbol{\delta}_1(\tau)'\boldsymbol{C}$$

The last term is asymptotically quadratic in $\boldsymbol{\delta}_1(\tau)$.

$$\begin{aligned}\mathbb{E}V_{TN}^{(4)}(\boldsymbol{\delta}_1(\tau)) &= 2\frac{\lambda_T}{T}\sum_{i=1}^{N}\int_0^{v_{it}}\sqrt{T}\left(G_i\left(s/\sqrt{T}\right)-G_i(0)\right)ds \\ &= \frac{\lambda_T}{T}\sum_{i=1}^{N}g_i(0)\left(\tilde{\boldsymbol{x}}_i'\boldsymbol{\delta}_1(\tau)/\sqrt{N}+\varphi_{it}-\left(\lambda_T/\sqrt{T}\right)\varphi_i\right)^2+o(1) \\ &\to \lambda\boldsymbol{\delta}_1(\tau)'\boldsymbol{D}_3\boldsymbol{\delta}_1(\tau)\end{aligned}$$

■

Proof of Corollary 1: From Theorem 1, the limiting form of the objective function is

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

thus its minimizer is,

$$\boldsymbol{\delta}_1(\tau) = (\boldsymbol{D}_1+2\lambda\boldsymbol{D}_3)^{-1}(\boldsymbol{B}+\lambda\boldsymbol{C})$$

which implies that $\mathbb{E}\boldsymbol{\delta}_1(\tau)=\boldsymbol{0}$, and

$$\mathrm{Var}(\boldsymbol{\delta}_1(\tau)) = (\boldsymbol{D}_1+2\lambda\boldsymbol{D}_3)^{-1}(\boldsymbol{D}_0+\lambda^2\boldsymbol{D}_2)(\boldsymbol{D}_1+2\lambda\boldsymbol{D}_3)^{-1}) = \boldsymbol{\Sigma}_1(\lambda)^{-1}\boldsymbol{\Sigma}_0(\lambda)\boldsymbol{\Sigma}_1(\lambda)^{-1}$$

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

Proof of Corollary 3: From Theorem 2, the limiting form of the objective function is

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

thus its minimizer is $\boldsymbol{\delta}_1=(\boldsymbol{H}_1+2\lambda\boldsymbol{H}_3)^{-1}(\boldsymbol{B}+\lambda\boldsymbol{C})$, which implies that $\mathbb{E}\boldsymbol{\delta}_1=\boldsymbol{0}$, and

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

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

Lemma 3. *Let $\mathcal{A}=[0,\bar{\lambda}]\subset\mathbb{R}_+$ where $\bar{\lambda}=12\zeta_d/\zeta_a+0.25\zeta_a-\epsilon>0$ for positive constants ζ_a, ζ_d, and ϵ. Then, the set $\mathcal{A}$ is non-empty, closed and bounded.*

Lemma 4. *Let $\mathcal{A}_{(n)}$ be a decreasing sequence of sets (e.g., $\mathcal{A}_{(1)}\supseteq\mathcal{A}_{(2)}\supseteq\ldots\supseteq\mathcal{A}_{(N)}$). 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 3, the set $\mathcal{D}$ is non-empty, closed and bounded. ■

Lemma 5. *Let (a) $\zeta_a,\zeta_b,\zeta_c,\zeta_d$ be positive constants, and (b) the parameter $\lambda\in\mathcal{A}$. Then, the rational function $\pi(\lambda):\mathcal{A}\to\mathbb{R}_+$,*

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

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

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

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

The second derivative is equal to,

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

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

Lemma 6. *If $\pi(\lambda)$ is a strictly convex function over $\mathcal{A} \subset \mathbb{R}_+$, the function $\pi(\lambda)$ is also strictly convex over $\mathcal{D} \subset \mathcal{A}$.*

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

Proof of Theorem 3: The asymptotic covariance matrix is equal to

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

Using simple matrix algebra, the matrix can be written as

$$\begin{pmatrix} \boldsymbol{\Sigma}_1(\lambda;\tau_1)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_1,\tau_1)\boldsymbol{\Sigma}_1(\lambda;\tau_1)^{-1} & \dots & \boldsymbol{\Sigma}_1(\lambda;\tau_1)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_1,\tau_J)\boldsymbol{\Sigma}_1(\lambda;\tau_J)^{-1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_1(\lambda;\tau_J)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_J,\tau_1)\boldsymbol{\Sigma}_1(\lambda;\tau_1)^{-1} & \dots & \boldsymbol{\Sigma}_1(\lambda;\tau_J)^{-1}\boldsymbol{\Sigma}_0(\lambda;\tau_J,\tau_J)\boldsymbol{\Sigma}_1(\lambda;\tau_J)^{-1} \end{pmatrix}$$

The trace of the matrix is,

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

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

Replacing the matrices in the last equation gives,

$$= \sum_{j=1}^{J} \mathrm{tr}\left\{(\boldsymbol{A}^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{B}^j)^{-1}\boldsymbol{C}^j(\boldsymbol{D}^j + 0.25\lambda^2\boldsymbol{I})(\boldsymbol{A}^j + 2\lambda\boldsymbol{I})^{-1}(\boldsymbol{B}^j)^{-1}\right\}$$

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

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

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

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

the last equation can be written as

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

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

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

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

$$\begin{aligned}&= \sum_{j=1}^{J}\mathrm{tr}\left\{(\boldsymbol{\Lambda}_a^j+2\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_b^j)^{-1}\boldsymbol{\Lambda}_c^j(\boldsymbol{\Lambda}_d^j+0.25\lambda^2\boldsymbol{I})(\boldsymbol{\Lambda}_a^j+2\lambda\boldsymbol{I})^{-1}(\boldsymbol{\Lambda}_b^j)^{-1}\right\} \\ &= \sum_{j=1}^{J}\sum_{i=1}^{p}\frac{\zeta_c^{ij}(\zeta_d^{ij}+0.25\lambda^2)}{(\zeta_b^{ij}(\zeta_a^{ij}+2\lambda))^2} = \sum_{j=1}^{J}\sum_{i=1}^{p}\pi(\lambda)^{ij} = \Pi(\lambda)\end{aligned}$$

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

Since $\boldsymbol{A}^j$, $\boldsymbol{C}^j$ and $\boldsymbol{D}^j$ are positive definite, their eigenvalues ζ_a^{ij}, ζ_c^{ij} and ζ_d^{ij} are positive for all i,j. Using Lemma 5, $\pi(\lambda)^{ij}$ is convex in λ in $\mathcal{A}^{ij}$ for all i,j. Since the sets $\mathcal{A}^{ij}$ are a decreasing sequence of sets, the functions $\pi(\lambda)^{ij}$ are also convex, using Lemma 6, in $\lambda\in\mathcal{D}=\cap_{ij}\mathcal{A}^{ij}$. Since the sum of convex functions is also convex, $\Pi(\lambda)$ is also convex in $\lambda\in\mathcal{D}$.

Therefore, the function $\Pi(\lambda):\mathcal{D}\to\mathbb{R}_+$ is a continuous strictly convex function defined on a non-empty, compact set (Lemma 4). These sufficient conditions implies that $\Pi(\lambda)$ has a unique minimizer, $\exists\lambda^*\in\mathcal{D}$ such that $\Pi(\lambda^*)<\Pi(\lambda)$ for all $\lambda\in\mathcal{D}$. ■

Proof of Corollary 5: The jkth diagonal element of the asymptotic covariance matrix is

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

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

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

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

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

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

$$\lambda^* = \frac{8A_{jk}D_{jk}}{C_{jk}B_{jk}} = \frac{8\Omega_{jj}\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\Psi}\tilde{\boldsymbol{x}}_{jk}}{\omega_j\boldsymbol{x}_k'\boldsymbol{M}_j'\boldsymbol{\Phi}_j\boldsymbol{M}_j\boldsymbol{x}_k\tilde{\boldsymbol{x}}_{jk}'\tilde{\boldsymbol{x}}_{jk}} \tag{A.3}$$

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

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

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

References

Abrevaya, J., and C. Dahl (2005): "The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Regression Estimation on Panel Data," Working Paper, Department of Economics, Purdue University.

Akaike, H. (1973): "Information Theory and An Extension of the Maximum Likelihood Principle," Second International Symposium on Information Theory.

Altonji, J. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," *Journal of Political Economy*, 94, S176–S215.

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.

Baltagi, B. (1981): "Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model," *Journal of Econometrics*, 17, 21–49.

——— (2001): *Econometric Analysis of Panel Data*. Wiley, New York, 2nd edn.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics: Methods and Applications*. Cambridge University Press.

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

Davison, A. C., and D. V. Hinkley (1997): *Bootstrap Methods and Their Applications*. Cambridge University Press.

Donoho, D., S. Chen, and M. Saunders (1998): "Atomic Decomposition by Basis Pursuit," *SIAM Journal of Scientific Computing*, 20, 33–61.

Graybill, F. (1969): *Introduction to Matrices with Applications in Statistics*. Wadsworth.

Hastie, T., R. Tibshirani, and H. Zou (2004): "On the "Degrees of Freedom" of the Lasso," mimeo, Department of Statistics, Stanford University.

Hendricks, W., and R. Koenker (1992): "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, 87, 58–68.

Hoerl, A., and R. Kennard (1988): "Ridge Regression," *Encyclopedia of Statistical Sciences*, 8, 129–136.

Horowitz, J., and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economics Studies*, 63, 145–168.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

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

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

——— (2006): "Quantreg," R package version 3.90, `www.r-project.org`.

Koenker, R., and G. Bassett (1982): "Tests of Linear Hypothesis and ℓ_1 Estimation," *Econometrica*, 50, 1577–1584.

Koenker, R., and K. Hallock (2000): "Quantile Regression: An Introduction," mimeo, Department of Economics, University of Illinois at Urbana-Champaign.

Koenker, R., and S. Portnoy (1987): "L-estimation for Linear Models," *Journal of the American Statistical Association*, 82, 851–857.

Kooperberg, C., and C. Stone (1991): "A Study of Logspline Density Estimation," *Computational Statistics and Data Analysis*, 12, 327–347.

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

Maddala, G. (1971): "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica*, 39, 343–358.

Nerlove, M. (1971): "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections," *Econometrica*, 39(2), 359–382.

Powell, J. L. (1991): "Estimation of Monotonic Regression Models under Quantile Restrictions," in *Nonparametric and Semiparametric Methods in Econometrics*, ed. by W. Barnett, J. Powell, and G. Tauchen. Cambridge University Press, Cambridge.

Rao, C. (1968): *Linear Statistical Inference and Its Applications*. Wiley, New York.

Robinson, G. K. (1991): "That BLUP Is a Good Thing: The Estimation of Random Effects (with discussion)," *Statistical Science*, 6(1), 15–51.

Ruppert, D., and R. Carroll (1980): "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828–838.

Schwartz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Scott, D. W. (1992): *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley and Sons, New York.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

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

Venables, W. N., and B. D. Ripley (2002): *Modern Applied Statistics with S*. Springer.

Ziliak, J. (1997): "Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business and Economic Statistics*, 15(4), 419–431.