

# Estimating and Testing a Quantile Regression Model with Interactive Effects\*

Matthew Harding<sup>†</sup> and Carlos Lamarche<sup>‡</sup>

November 16, 2011

---

## Abstract

This paper proposes a quantile regression estimator for a panel data model with interactive effects potentially correlated with the independent variables. We provide conditions under which the slope parameter estimator is asymptotically Gaussian. Monte Carlo studies are carried out to investigate the finite sample performance of the proposed method in comparison with other candidate methods. We discuss an approach to testing the model specification against a competing fixed effects specification. The paper presents an empirical application of the method to study the effect of class size and class composition on educational attainment. The findings show that (i) a change in the gender composition of a class impacts differently low- and high-performing students; (ii) while smaller classes are beneficial for low performers, larger classes are beneficial for high performers; (iii) reductions in class size do not seem to impact mean and median student performance; (iv) the fixed effects specification is rejected in favor of the interactive effects specification.

*JEL: C23, C33*

*Keywords: Quantile Regression; Panel data; Interactive effects; Instrumental variables.*

---

---

\*We are grateful to Giacomo De Giorgi, Antonio Galvao, Jerry Hausman, Caroline Hoxby, Roger Koenker, Whitney Newey and seminar participants at Stanford University, the University of Oklahoma, the International Symposium on Econometrics of Specification Tests in 30 Years, California Econometrics Conference 2010, New York Camp Econometrics V, and the 16th International Conference on Panel Data for useful comments. We thank Michele Pellizzari for providing the data for the empirical section. The R software for the method introduced in this paper (as well as the other methods discussed in this paper) are available upon request and for download from the authors' websites.

<sup>†</sup>Corresponding author: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: [mch@stanford.edu](mailto:mch@stanford.edu)

<sup>‡</sup>Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019; Phone: (405) 325-5857; Email: [lamarche@ou.edu](mailto:lamarche@ou.edu)

## 1. Introduction

Panel data models which account for the confounding effect of unobservable individual effects have become the models of choice in many applied areas of economics from microeconomics to finance. Recent papers have focused on relaxing the traditional fixed effects framework by allowing for multiple interactive effects (Bai, 2009; Pesaran, 2006). The natural extension of the classical panel data models with  $N$  cross-sectional units and  $T$  time periods (Hsiao 2003, Baltagi 2008) is thus  $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t + u_{it}$ , where  $\boldsymbol{\lambda}_i$  is an  $r \times 1$  vector of factor loadings and  $\mathbf{f}_t$  corresponds to the  $r$  common time-varying factors, and where both  $\boldsymbol{\lambda}_i$  and  $\mathbf{f}_t$  are latent variables. Although this extension substantially increases the flexibility of controlling for unobserved heterogeneity, the existing estimation approaches are designed for Gaussian models and do not offer the possibility of estimating heterogeneous covariate effects, which may be of interest to applied researchers. For example, Bandiera, Larcinese, and Rasul (2010) argue for the use of heterogeneous effects in the design of educational policies.

This paper proposes a panel data quantile regression estimator for a model with interactive effects, allowing  $\boldsymbol{\lambda}$  and  $\mathbf{f}$  to be correlated with the independent variables. We also allow for the possibility that the covariate  $\mathbf{x}$  is stochastically dependent on  $u$ . We introduce a panel data version of the instrumental variable estimator proposed by Chernozhukov and Hansen (2005, 2006, 2008), while at the same time accounting for latent heterogeneity. Our method differs from Harding and Lamarche (2009) and Galvao (2009) because it does not consider the case of unobserved heterogeneity represented by a classical individual effect  $\lambda_i$ . We provide conditions under which the slope parameter estimator is consistent and asymptotically Gaussian. Moreover, we investigate the finite sample performance of the proposed method in comparison to other candidate methods. Monte Carlo evidence shows that the finite sample performance of the proposed method is satisfactory in all the variants of the models, including specifications with  $\boldsymbol{\lambda}$ ,  $\mathbf{f}$  and  $u$  correlated with the independent variable  $\mathbf{x}$ . While the estimation of nuisance parameters in a large  $N$  panel data quantile regression model may be regarded by applied researchers as computationally demanding, this paper solves a relatively simple linear programming problem that performs extremely well in large size applications.

We apply the approach to reexamine an often controversial topic in the social sciences, estimating the distributional effect of class size and class composition on educational attainment, using a unique dataset of an exogenous allocation of students into classes at Bocconi University (De Giorgi,

Pellizzari and Woolston 2009). We estimate quantile treatment effects, while relaxing the assumption that the individual latent variables are class invariant. If students' motivation and teachers' quality enter multiplicatively in the educational attainment function, standard approaches would produce biased results. Therefore, we estimate a model that allows for the possibility that a teacher's quality affects performance only if the student is motivated and receptive to instruction. We find that the proposed method gives different policy prescriptions relative to standard methods. While a reduction in class size does not impact mean and median student performance, it affects performance at the tails of the conditional distribution. Our finding suggests that this policy benefits weak students, but harms high achievers. Moreover, we find evidence that indicates that a change in the gender composition of a class impacts differently low- and high-performing students.

Our paper complements the recent focus on heterogeneous treatment effects in the applied econometrics literature (DiNardo and Lee, 2010). In most applications with endogenous right hand side variables, such as a treatment indicator, it is also particularly informative to consider the possibility of heterogeneous treatment effects (Heckman and Vytlacil, 2001). Quantile regression provides a convenient way to introduce a type of heterogeneous treatment effect (e.g., Lehmann 1974, Doksum 1974, Koenker 2005) across individuals conditional on the quantile of the outcome distribution. The literature investigating quantile regression estimation of the classical static panel data model is still relatively new. While Koenker (2004) introduces a class of penalized quantile regression estimators, Lamarche (2010) provides conditions under which it is possible to obtain the minimum variance estimator in the class of penalized estimators, the analog of the GLS in the class of penalized least squares estimators for panel data. Abrevaya and Dahl (2008) consider the classical correlated random effects model and Harding and Lamarche (2009) estimate a model with endogenous covariates. Our paper is also related to Galvao (2009), who proposes an instrumental variable approach for estimating a dynamic panel data model. Alternative models and approaches to the one considered in this paper are introduced by Canay (2010), Chernozhukov, Fernández-Val, and Newey (2009), Powell (2009), Wei and He (2006), Ando and Tsay (2010), Rosen (2009), and Ponomareva (2010). The analysis of an incidental parameter problem in quantile regression with fixed effects is described in Graham, Hahn, and Powell (2009) and Kato and Galvao (2010).

The next section presents the basic idea, the model and an estimator. Section 3 introduces the quantile regression approach and Section 4 studies the asymptotic properties of the estimator. Section 5 offers Monte-Carlo evidence. Section 6 demonstrates how the estimator can be used in an empirical application to the estimation of class size effects for university students. Section 7 concludes.

## 2. An Estimation Approach

Consider the following model:

$$(2.1) \quad y_{it} = \boldsymbol{\alpha}' \mathbf{d}_{it} + \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\lambda}'_i \mathbf{f}_t + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T$$

$$(2.2) \quad \mathbf{d}_{it} = \boldsymbol{\Pi}'_1 \mathbf{w}_{it} + \boldsymbol{\Pi}'_2 \mathbf{x}_{it} + \boldsymbol{\Pi}'_3 \mathbf{f}_t + \boldsymbol{\Pi}'_4 \boldsymbol{\lambda}'_i \mathbf{f}_t + \boldsymbol{\Pi}'_5 \boldsymbol{\lambda}_i + \mathbf{v}_{it}$$

The first equation is a panel data model with interactive effects. The variable  $y_{it}$  is the response for subject  $i$  at time  $t$ ,  $\mathbf{d}$  is a vector of  $k_1$  endogenous variables,  $\mathbf{x}$  is a vector of  $k_2$  exogenous independent variables,  $\boldsymbol{\lambda}_i$  is a vector of  $r$  unobserved loadings,  $\mathbf{f}_t$  is a vector of  $r$  latent factors, and  $u$  is the error term. The parameter of interest is  $\boldsymbol{\alpha}$ , while the interactive effects  $\boldsymbol{\lambda}'_i \mathbf{f}_t$  are treated as nuisance parameters. The second equation indicates that  $\mathbf{d}$  is correlated with a vector of  $m \geq k_1$  instruments  $\mathbf{w}$ , the exogenous variables  $\mathbf{x}$ , and the interactive effects  $\boldsymbol{\lambda}_i$  and  $\mathbf{f}_t$ . We assume that the variable  $\mathbf{v}$  is stochastically dependent on  $u$ . It is convenient to write equation 2.1 in a more concise matrix notation,

$$(2.3) \quad \mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\lambda} + \mathbf{u}$$

where  $\mathbf{y}$  is an  $NT \times 1$  vector,  $\mathbf{D}$  is an  $NT \times k_1$  matrix,  $\mathbf{X}$  is an  $NT \times k_2$  matrix, and  $\mathbf{F}$  is an  $NT \times r$  matrix. We let  $\mathbf{W}$  be a matrix of instruments of dimension  $NT \times m$ .

The recent panel literature with  $T$  large and  $N$  large develops least squares estimation procedures for model 2.1 and 2.2 under the assumption that  $u$  and  $\mathbf{v}$  are independent random variables (see, e.g., Pesaran 2006, Bai 2009). Considering the following conditions, we tentatively propose an estimation approach that allows for dependence between  $u$  and  $\mathbf{v}$ .

**ASSUMPTION 1.**  $(u_{it}, \mathbf{v}'_{it})'$  satisfies  $(u_{it}, \mathbf{v}'_{it})' = \sum_{l=0}^{\infty} \boldsymbol{\alpha}_{il} \boldsymbol{\zeta}_{i,t-l}$ , where  $\boldsymbol{\zeta}_{it}$  is a vector of identically, independently distributed (IID) random variables with mean zero, variance matrix  $\mathbf{I}_{k_1+1}$ , and finite fourth order cumulants. In particular  $\text{Var}((u_{it}, \mathbf{v}'_{it})') = \boldsymbol{\Sigma} < \infty$  for all  $i, t$ , for some constant positive definite matrix  $\boldsymbol{\Sigma}$ .

**ASSUMPTION 2.** The  $r \times 1$  vector  $\mathbf{f}_t$  is drawn from a zero mean, unit variance, covariance stationary process, with absolute summable autocovariances, distributed independently of  $u_{it}$  and  $\mathbf{v}'_{it}$  for all  $i, t, t'$ .

**ASSUMPTION 3.** The factor loadings  $\boldsymbol{\lambda}_i = \boldsymbol{\lambda} + \boldsymbol{\pi}_i$  are distributed independently of  $u_{jt}$  and  $\mathbf{v}_{jt}$  for all  $i$  and  $j$  with mean  $\boldsymbol{\lambda}$  and finite variances.

**ASSUMPTION 4.** The variables  $\mathbf{w}_{it}$  and  $u_{it}$  are stochastically independent and the number of endogenous variables  $k_1$  is equal to the number of instruments  $m$ .

There is a natural connection between econometric approaches to factor models with latent variables and instrumental variable (IV) methods. While the IV estimation of high-dimensional factor models could be problematic in large datasets, its use may provide a solution in the case of  $T$  small and  $N$  large. To illustrate the approach, we consider,

$$(2.4) \quad \mathbf{y}(\boldsymbol{\alpha}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\lambda} + \mathbf{W}\boldsymbol{\eta} + \mathbf{u},$$

where  $\mathbf{y}(\boldsymbol{\alpha}) = \mathbf{y} - \mathbf{D}\boldsymbol{\alpha}$ . Then, the estimator  $\hat{\boldsymbol{\eta}}(\boldsymbol{\alpha}) = (\mathbf{W}'\mathbf{M}\mathbf{W})^{-1}\mathbf{W}'\mathbf{M}\mathbf{y}(\boldsymbol{\alpha})$ , where  $\mathbf{M} = \mathbf{I} - \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$  and  $\mathbf{H} = [\mathbf{X}:\mathbf{F}]$ . The (infeasible) instrumental variable estimator  $\hat{\boldsymbol{\alpha}}$  of the parameter of interest  $\boldsymbol{\alpha}$  is defined as,

$$(2.5) \quad \hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha} \in \mathcal{A}}{\operatorname{argmin}} \{ \hat{\boldsymbol{\eta}}(\boldsymbol{\alpha})' \mathbf{W}' \mathbf{M} \mathbf{W} \hat{\boldsymbol{\eta}}(\boldsymbol{\alpha}) \}$$

In practice, the projection matrix  $\mathbf{M}$  can be constructed by following Pesaran's (2006) method. Substituting (2.2) into (2.1), combining the two equations, and summing over the cross-sectional dimension of the model, we obtain,

$$(2.6) \quad \bar{\mathbf{Z}} = \bar{\mathbf{W}}\mathbf{C}'_1 + \bar{\mathbf{X}}\mathbf{C}'_2 + \mathbf{F}(\bar{\mathbf{C}}_3 + \bar{\boldsymbol{\lambda}}\mathbf{C}'_4) + \bar{\boldsymbol{\Lambda}}\mathbf{C}'_5 + \bar{\boldsymbol{\Xi}},$$

where  $\bar{\mathbf{C}}_3 = N^{-1} \sum_{i=1}^N ((\boldsymbol{\Pi}_{3i}\boldsymbol{\alpha})', \boldsymbol{\Pi}'_{3i})'$ ,  $\bar{\boldsymbol{\lambda}} = N^{-1} \sum_{i=1}^N \boldsymbol{\lambda}_i$  and,

$$\underbrace{\mathbf{C}_1}_{(1+k_1) \times m} = \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Pi}'_1 \\ \boldsymbol{\Pi}'_1 \end{pmatrix}; \quad \underbrace{\mathbf{C}_2}_{(1+k_1) \times k_2} = \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Pi}'_2 + \boldsymbol{\beta}' \\ \boldsymbol{\Pi}'_2 \end{pmatrix}; \quad \underbrace{\mathbf{C}_4}_{(1+k_1) \times 1} = \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Pi}_4 + 1 \\ \boldsymbol{\Pi}_4 \end{pmatrix}; \quad \underbrace{\mathbf{C}_5}_{(1+k_1) \times r} = \begin{pmatrix} \boldsymbol{\alpha}'\boldsymbol{\Pi}_5 \\ \boldsymbol{\Pi}_5 \end{pmatrix},$$

The matrix  $\bar{\mathbf{Z}} = \bar{\mathbf{z}} \otimes \iota_N$ ,  $\bar{\mathbf{z}} = (\bar{\mathbf{z}}_1, \dots, \bar{\mathbf{z}}_T)'$ , with  $\mathbf{z}_t = ((y_{1t}, \mathbf{d}'_{1t}), \dots, (y_{Nt}, \mathbf{d}'_{Nt}))'$ . The matrices  $\bar{\mathbf{W}} = \bar{\mathbf{w}} \otimes \iota_N$ ,  $\bar{\mathbf{X}} = \bar{\mathbf{x}} \otimes \iota_N$ ,  $\bar{\boldsymbol{\Lambda}} = \bar{\boldsymbol{\lambda}} \otimes \iota_{TN}$  are similarly defined. As usual,  $\otimes$  denotes Kronecker product and  $\iota_s$  a vector  $(1, 1, \dots, 1)' \in \mathbb{R}^s$ .

**ASSUMPTION 5.** *The matrix  $\bar{\mathbf{C}}_3 + \mathbf{C}_4\bar{\boldsymbol{\lambda}}'$  converges to a limiting matrix  $\tilde{\mathbf{C}}_3$  with rank  $k_1 + 1 < r$ .*

Moreover, the error term in equation 2.6 is a matrix  $\bar{\boldsymbol{\Xi}} = \bar{\boldsymbol{\xi}} \otimes \iota_N$  of dimension  $NT \times (k_1 + 1)$ , with  $\bar{\boldsymbol{\xi}} = (\bar{\boldsymbol{\xi}}_1, \dots, \bar{\boldsymbol{\xi}}_T)'$  and  $\bar{\boldsymbol{\xi}}_t = N^{-1} \sum_{i=1}^N (\boldsymbol{\alpha}'\mathbf{v}_{it} + u_{it}, \mathbf{v}_{it})'$ . Applying Lemma 1 in Pesaran (2006) under condition 1, it is possible to show that each coordinate on the vector  $\bar{\boldsymbol{\xi}}$ , say  $\bar{\boldsymbol{\xi}}_t$ , converges to zero in probability. Letting  $\boldsymbol{\Gamma}_0 = \tilde{\mathbf{C}}_3(\tilde{\mathbf{C}}_3'\tilde{\mathbf{C}}_3)^{-1}$ ,  $\boldsymbol{\Gamma}_1 = \mathbf{C}'_1\boldsymbol{\Gamma}_0$ ,  $\boldsymbol{\Gamma}_2 = \mathbf{C}'_2\boldsymbol{\Gamma}_0$ , and  $\boldsymbol{\Gamma}_5 = \mathbf{C}'_5\boldsymbol{\Gamma}_0$ , we have that,

$$(2.7) \quad \mathbf{F} = \bar{\mathbf{Z}}\boldsymbol{\Gamma}_0 - \bar{\mathbf{W}}\boldsymbol{\Gamma}_1 - \bar{\mathbf{X}}\boldsymbol{\Gamma}_2 - \bar{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}_5,$$

which suggest that the unknown factors can be approximated by  $(\bar{\mathbf{z}}'_t, \bar{\mathbf{x}}'_t, \bar{\mathbf{w}}'_t, 1)'$ . We note however that due to the lack of independence between  $\mathbf{d}_{it}$  and  $u_{it}$ , the vector  $\bar{\mathbf{z}}_t$  may include endogenous covariates. Using a simple reparametrization, we obtain,

$$(2.8) \quad \mathbf{F}\boldsymbol{\lambda} = \bar{\mathbf{D}}\boldsymbol{\delta}_0 + \bar{\mathbf{Y}}\boldsymbol{\delta}_1 + \bar{\mathbf{W}}\boldsymbol{\delta}_2 + \bar{\mathbf{X}}\boldsymbol{\delta}_3 + \boldsymbol{\delta}_4.$$

Replacing 2.8 in equation 2.4, we obtain,

$$(2.9) \quad \mathbf{y}(\boldsymbol{\alpha}, \boldsymbol{\delta}_0) = \tilde{\mathbf{H}}\boldsymbol{\delta} + \tilde{\mathbf{W}}\boldsymbol{\gamma} + \mathbf{u}$$

where  $\mathbf{y}(\boldsymbol{\alpha}, \boldsymbol{\delta}_0) = \mathbf{y}(\boldsymbol{\alpha}) - \bar{\mathbf{D}}\boldsymbol{\delta}_0$ ,  $\tilde{\mathbf{H}} = [\mathbf{X}, \bar{\mathbf{Y}}, \bar{\mathbf{X}}, \iota_{NT}]$ ,  $\tilde{\mathbf{W}} = [\mathbf{W}, \bar{\mathbf{W}}]$ ,  $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2, \boldsymbol{\delta}'_4)'$ , and  $\boldsymbol{\gamma} = (\boldsymbol{\eta}', \boldsymbol{\delta}'_2)'$ . Letting  $\boldsymbol{\vartheta} = (\boldsymbol{\alpha}', \boldsymbol{\delta}'_0)'$ , we have that,

$$(2.10) \quad \hat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta} \in \Theta}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\gamma}}(\boldsymbol{\vartheta})' \tilde{\mathbf{W}}' \tilde{\mathbf{M}} \tilde{\mathbf{W}} \hat{\boldsymbol{\gamma}}(\boldsymbol{\vartheta}) \right\}$$

where  $\hat{\boldsymbol{\gamma}}(\boldsymbol{\vartheta}) = (\tilde{\mathbf{W}}' \tilde{\mathbf{M}} \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \tilde{\mathbf{M}} \mathbf{y}(\boldsymbol{\vartheta})$  and  $\tilde{\mathbf{M}} = \mathbf{I} - \tilde{\mathbf{H}}(\tilde{\mathbf{H}}' \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}'$ . The estimator of  $\boldsymbol{\vartheta}$  is equal to  $\hat{\boldsymbol{\vartheta}} = (\tilde{\mathbf{D}}' \tilde{\mathbf{P}} \tilde{\mathbf{D}})^{-1} (\tilde{\mathbf{D}}' \tilde{\mathbf{P}} \mathbf{y})$ , where  $\tilde{\mathbf{P}} = \tilde{\mathbf{M}} \tilde{\mathbf{W}} (\tilde{\mathbf{W}}' \tilde{\mathbf{M}} \tilde{\mathbf{W}})^{-1} \tilde{\mathbf{W}}' \tilde{\mathbf{M}}$ , and  $\tilde{\mathbf{D}} = [\mathbf{D}, \bar{\mathbf{D}}]$ . The following proposition shows that if valid instruments  $\mathbf{w}_{it}$  are present, we can perform an instrumental variable regression augmented by the exogenous right-hand side variables in equation 2.7, leading to consistent estimates of the parameter of interest.

**PROPOSITION 1.** *Under the previous conditions, a feasible and consistent estimator of  $\boldsymbol{\alpha}$  is equal to  $\hat{\boldsymbol{\alpha}} = (\mathbf{D}' \tilde{\mathbf{M}}_{\tilde{\mathbf{P}}} \mathbf{D})^{-1} (\mathbf{D}' \tilde{\mathbf{M}}_{\tilde{\mathbf{P}}} \mathbf{y})$ , where  $\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}} = \tilde{\mathbf{P}} - \tilde{\mathbf{P}} \bar{\mathbf{D}} (\bar{\mathbf{D}}' \tilde{\mathbf{P}} \bar{\mathbf{D}})^{-1} \bar{\mathbf{D}}' \tilde{\mathbf{P}}$ .*

Proposition 1 indicates that it is possible to obtain a feasible estimator for the slope parameter of interest  $\boldsymbol{\alpha}$  in a model with interactive effects and endogenous covariates. The method extends Pesaran's (2006) analysis accommodating to issues associated with the lack of independence between  $\mathbf{d}$  and  $(\boldsymbol{\lambda}', \mathbf{u})'$ . The next Section shows that the same strategy can be employed to estimate a quantile regression model with interactive effects and endogenous covariates.

### 3. A Quantile Regression Approach

We being considering for simplicity the following conditional quantile functions:

$$(3.1) \quad Q_{Y_{it}}(\tau | \mathbf{d}_{it}, \mathbf{x}_{it}, \boldsymbol{\lambda}_i, \mathbf{f}_t) = \boldsymbol{\alpha}' \mathbf{d}_{it} + \boldsymbol{\beta}' \mathbf{x}_{it} + \boldsymbol{\lambda}'_i \mathbf{f}_t + G_u(\tau)^{-1},$$

$$(3.2) \quad Q_{D_{it}}(\tau | \mathbf{w}_{it}, \mathbf{x}_{it}, \boldsymbol{\lambda}_i, \mathbf{f}_t) = \boldsymbol{\Pi}'_1 \mathbf{w}_{it} + \boldsymbol{\Pi}'_2 \mathbf{x}_{it} + \boldsymbol{\Pi}'_{3i} \mathbf{f}_t + \boldsymbol{\Pi}'_4 \boldsymbol{\lambda}'_i \mathbf{f}_t + \boldsymbol{\Pi}'_5 \boldsymbol{\lambda}_i + \kappa G_u(\tau)^{-1},$$

where  $\tau$  is a quantile in the interval  $(0, 1)$ ,  $\kappa$  is a parameter, and  $G$  denotes the distribution function of the *iid* error term  $u$ . This model 3.1-3.2 is the simplest quantile regression version of model 2.1-2.2. A natural generalization of the model is defining the quantile function 3.2 for  $G_v(\tau')^{-1}$ . To estimate this model, we can accommodate the instrumental variable approach proposed in Chernozhukov and Hansen (2005) to panel data, integrating out the quantile  $\tau'$  as in Ma and Koenker (2006). If we augment the design matrix with cross-sectional averages, we have that the approximation for the factors  $\mathbf{f}$ 's depend on the quantiles  $\tau$  and  $\tau'$ . By integrating out the quantile

$\tau'$ , we may define the factors in terms of  $\tau$ . Chernozhukov and Hansen (2005) illustrate that their approach is always applicable to a class of triangular models.

The model can be easily generalized to the standard location-scale shift model and other more general panel data models with heterogeneous effects. As it will be clear in Section 6, we will use the general version of these equations as a model for educational achievement. Following the literature (see, e.g., Ma and Koenker 2006, Hanushek et al. 2003, De Giorgi, Pellizzari and Woolston 2009, Bandiera, Larcinese, and Rasul 2010), the response variable  $y$  is educational attainment and is influenced by class size and peer effects  $\mathbf{d}$ , and individual, family and school characteristics  $\mathbf{x}$ . The last term  $u$  may represent idiosyncratic shocks to achievement which force the student to switch class.

A central concern in the estimation of the distributional effects of class size is unobserved heterogeneity. While most of the models estimated in the literature assume the classical additive separable structure on unobserved heterogeneity  $\lambda + f$ , we will estimate a more general specification allowing for interactive effects. The variable  $\boldsymbol{\lambda}$  captures a student's unobserved ability to absorb knowledge when listening to lectures, effort and motivation, and the variable  $\mathbf{f}$  includes teachers' quality and other class-invariant unobserved effects. The education production function 3.1 incorporates unobserved heterogeneity, while allowing for the possibility that teachers' quality affects performance *only if* the student is motivated and receptive to instruction.

Proposition 1 suggests a strategy that can be employed to estimate a quantile regression model with interactive effects and endogenous covariates. We define,

$$(3.3) \quad C_{it}(\tau, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}) = \rho_{\tau} \left( y_{it} - \mathbf{d}'_{it}\boldsymbol{\alpha} - \mathbf{x}'_{it}\boldsymbol{\beta} - \hat{\mathbf{f}}'_t(\tau)\boldsymbol{\delta} - \hat{\boldsymbol{\Phi}}'_{it}(\tau)\boldsymbol{\gamma} \right).$$

where  $\rho_{\tau}(u) = u(\tau - I(u \leq 0))$  is the standard quantile loss function (see, e.g., Koenker 2005). The quantile regression check function includes two additional terms that deserve our attention. Using the convention that the conditional quantile function  $Q_{Y_{it}}(\tau|\mathbf{d}_{it}, \mathbf{x}_{it}, \boldsymbol{\lambda}_i, \mathbf{f}_t)$  is evaluated at  $\mathbf{d}_{it} = Q_{D_{it}}(\tau|\mathbf{w}_{it}, \mathbf{x}_{it}, \boldsymbol{\lambda}_i, \mathbf{f}_t)$ , we can substitute (3.2) into (3.1) and summing over the cross-sectional dimension of the model, we obtain,

$$(3.4) \quad \bar{\mathbf{z}}_t(\tau) = \mathbf{C}_1\bar{\mathbf{w}}_t + \mathbf{C}_2(\tau)\bar{\mathbf{x}}_t + (\bar{\mathbf{C}}_3 + \mathbf{C}_4\bar{\boldsymbol{\lambda}}')\mathbf{f}_t + \mathbf{C}_5\bar{\boldsymbol{\lambda}},$$

where  $\bar{\mathbf{z}}_t(\tau)$  is the cross-sectional average of  $\mathbf{z}_{it}(\tau) = (Q_{Y_{it}}(\tau|\bullet), Q_{D_{it}}(\tau|\bullet))'$ , and  $\mathbf{C}_2(\tau) = ((\boldsymbol{\alpha}'\boldsymbol{\Pi}'_2 + \boldsymbol{\beta}(\tau)')', \boldsymbol{\Pi}'_2)'$ . As before we can augment the design matrix by adding the cross-sectional averages of the endogenous and exogenous variables,  $\mathbf{f}_t(\tau) = \boldsymbol{\Psi}(\tau; \bar{\mathbf{z}}_t, \bar{\mathbf{w}}_t, \bar{\mathbf{x}}_t, 1)$ . The second term  $\boldsymbol{\Phi}_{it}(\tau) = \boldsymbol{\Phi}(\tau; \mathbf{w}_{it}, \mathbf{x}_{it}, \mathbf{f}_t, \boldsymbol{\lambda}_i)$  is a vector of transformations of instruments. It is possible to estimate  $\boldsymbol{\Phi}$

by a least squares projection of the endogenous variables  $\mathbf{d}$  on the instruments  $\mathbf{w}$ , the exogenous variables  $\mathbf{x}$ , and a vector of individual and time effects. We consider the case of  $\dim(\boldsymbol{\alpha}) = \dim(\boldsymbol{\gamma})$ , although the vector  $\boldsymbol{\Phi}$  may include more elements than the vector  $\mathbf{d}$ .

**Remark 1.** If cross-sectional averages are used instead of the conditional quantiles  $Q_D$ 's, the function  $C_{it}(\tau, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma})$  in 3.3 is replaced by,

$$(3.5) \quad C_{it}(\tau, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}) = \rho_\tau \left( y_{it} - \tilde{\mathbf{d}}_{it}' \boldsymbol{\theta} - \mathbf{x}'_{it} \boldsymbol{\beta} - \hat{\mathbf{f}}'_t(\tau) \boldsymbol{\delta} - \hat{\boldsymbol{\Phi}}'_{it}(\tau) \boldsymbol{\gamma} \right),$$

where  $\tilde{\mathbf{d}}_{it} = (\mathbf{d}'_{it}, \bar{\mathbf{d}}'_t)'$ . In this case,  $\mathbf{f}_t(\tau)$  is defined as  $\Psi(\tau; \bar{y}_t, \bar{\mathbf{x}}_t, 1)$  and  $\boldsymbol{\Phi}_{it}(\tau)$  is defined as  $\boldsymbol{\Phi}(\tau; \mathbf{w}_{it}, \bar{\mathbf{w}}_t)$ . Therefore, as in Proposition 1, we may instrument  $\tilde{\mathbf{d}}_{it}$  by the vector of instruments  $\tilde{\mathbf{w}}_{it} = (\mathbf{w}'_{it}, \bar{\mathbf{w}}'_t)'$ .

The procedure is similar in spirit to Chernozhukov and Hansen (2006) applied to panel models as in Harding and Lamarche (2009) and Galvao (2009). First, we minimize the objective function above for  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$ , and  $\boldsymbol{\delta}$  as functions of  $\tau$  and  $\boldsymbol{\alpha}$ ,

$$(3.6) \quad \{\hat{\boldsymbol{\beta}}(\tau, \boldsymbol{\alpha}), \hat{\boldsymbol{\delta}}(\tau, \boldsymbol{\alpha}), \hat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha})\} = \underset{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta} \in \mathcal{B} \times \mathcal{G} \times \mathcal{F}}{\operatorname{argmin}} \sum_{t=1}^T \sum_{i=1}^N C_{it}(\tau, \boldsymbol{\alpha}; \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}).$$

Then we estimate the coefficient on the endogenous variable by finding the value of  $\boldsymbol{\alpha}$ , which minimizes a weighted distance function defined on  $\boldsymbol{\gamma}$ :

$$(3.7) \quad \hat{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha} \in \mathcal{A}}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha})' \hat{\mathbf{A}}(\tau) \hat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha}) \right\}$$

for a positive definite matrix  $\mathbf{A}$ . Then, the quantile regression estimator for a model with interactive effects (QRIE) is defined as,

$$\hat{\boldsymbol{\theta}}(\tau) \equiv \left( \hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau), \hat{\boldsymbol{\delta}}(\tau) \right) = \left( \hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}(\tau), \tau), \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\alpha}}(\tau), \tau) \right).$$

It is straightforward to accommodate our estimator to the case of individual location shifts considered in Koenker (2004). We define,

$$(3.8) \quad C_{it}(\tau, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}, \lambda_i) = \rho_\tau \left( y_{it} - \mathbf{d}'_{it} \boldsymbol{\alpha} - \mathbf{x}'_{it} \boldsymbol{\beta} - \lambda_i - \hat{\mathbf{f}}'_t(\tau) \boldsymbol{\delta} - \hat{\boldsymbol{\Phi}}'_{it}(\tau) \boldsymbol{\gamma} \right).$$

where  $\rho_\tau(u) = u(\tau - I(u \leq 0))$  is the standard quantile loss function and  $\lambda_i$  is an individual effect. First, we minimize the objective function above for  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\delta}$  and  $\boldsymbol{\lambda}$  as functions of  $\tau$  and  $\boldsymbol{\alpha}$ ,

$$(3.9) \quad \{\tilde{\boldsymbol{\beta}}(\tau, \boldsymbol{\alpha}), \tilde{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha}), \tilde{\boldsymbol{\lambda}}(\tau, \boldsymbol{\alpha}), \tilde{\boldsymbol{\delta}}(\tau, \boldsymbol{\alpha})\} = \underset{\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\lambda} \in \mathcal{B} \times \mathcal{G} \times \mathcal{F} \times \Lambda}{\operatorname{argmin}} \sum_{j=1}^J \sum_{t=1}^T \sum_{i=1}^N C_{it}(\tau_j, \boldsymbol{\alpha}; \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\gamma}).$$



Then, we estimate the coefficient on the endogenous variable by:

$$(3.10) \quad \tilde{\alpha}(\tau) = \underset{\alpha \in \mathcal{A}}{\operatorname{argmin}} \left\{ \tilde{\gamma}(\tau, \alpha)' \tilde{\mathbf{A}}(\tau) \tilde{\gamma}(\tau, \alpha) \right\}$$

for a positive definite matrix  $\tilde{\mathbf{A}}(\tau)$ . The quantile regression estimator for a model with individual and interactive effects (QRIFE) is,

$$\tilde{\boldsymbol{\theta}}(\tau) \equiv \left( \tilde{\alpha}(\tau), \tilde{\boldsymbol{\beta}}(\tau), \tilde{\boldsymbol{\lambda}}(\tau), \tilde{\boldsymbol{\delta}}(\tau) \right) = \left( \tilde{\alpha}(\tau), \tilde{\boldsymbol{\beta}}(\tilde{\alpha}(\tau), \tau), \tilde{\boldsymbol{\lambda}}(\tilde{\alpha}(\tau), \tau), \tilde{\boldsymbol{\delta}}(\tilde{\alpha}(\tau), \tau) \right).$$

The following examples are designed to establish a connection between our methods and existing approaches in the literature.

**Example 1.** Consider the case that  $f_t = r = 1$  for all  $t$ . For simplicity, we let  $\boldsymbol{\Pi}_{3i} = \mathbf{0}$  in equation 3.2 and  $\tilde{\mathbf{C}}_4 = \mathbf{C}_4 + \mathbf{C}_5$ . From equation 3.4, we have that,

$$(3.11) \quad \tilde{\mathbf{C}}_4 \lambda_i = \mathbf{z}_{it}(\tau) - \mathbf{C}_1 \mathbf{w}_{it} - \mathbf{C}_2(\tau) \mathbf{x}_{it}$$

or, similarly, by taking averages over the time-series dimension of the model,

$$(3.12) \quad \lambda_i = \boldsymbol{\Psi}(\tau)' \bar{\mathbf{s}}_i,$$

where  $\mathbf{s}_{it} = (\mathbf{z}_{it}(\tau)', \mathbf{w}'_{it}, \mathbf{x}'_{it})'$ ,  $\bar{\mathbf{s}}_i$  denotes the sample average of  $\mathbf{s}$ , and  $\boldsymbol{\Psi}(\tau)$  is a vector of coefficients that includes  $\tilde{\mathbf{C}}_4$ ,  $\mathbf{C}_1$ , and  $\mathbf{C}_2$ . Therefore, in the one-factor model with fixed effects  $\lambda_i$ 's, we can augment the model by a vector of observables  $\bar{\mathbf{s}}_i$ . In the linear least squares case, it is known that the classical fixed effects estimator is numerically equivalent to the estimator obtained by augmenting the model by a vector of observables (Chamberlain 1982, Mundlak 1978).

**Example 2.** In the simplest conceivable case of  $\mathbf{f}_t$  not depending on the quantile  $\tau$  and  $u_{it}$  stochastically independent on  $\mathbf{v}_{it}$ , one can interpret that the method considers proxing  $\mathbf{f}_t$  with  $\bar{\mathbf{s}}_t = (\bar{\mathbf{z}}'_t, \bar{\mathbf{x}}'_t)'$ , implicitly assuming that  $\mathbf{f}_t = \boldsymbol{\Pi}' \bar{\mathbf{s}}_t + \boldsymbol{\epsilon}_t$ . To address the possibility that the loadings  $\lambda_i$ 's are correlated with the independent variables, Bai (2009) notices that one can similarly write  $\lambda_i = \boldsymbol{\Psi}' \bar{\mathbf{s}}_i + \boldsymbol{\eta}_i$ , where  $\bar{\mathbf{s}}_i = (\bar{\mathbf{d}}'_i, \bar{\mathbf{x}}'_i)'$ . The errors  $\boldsymbol{\epsilon}_t$  and  $\boldsymbol{\eta}_i$  are assumed to be distributed as  $F_\epsilon$  and  $F_\eta$ , and are independent of the covariates. Notice that,

$$(3.13) \quad \lambda'_i \mathbf{f}_t = (\boldsymbol{\Psi} \bar{\mathbf{s}}_i + \boldsymbol{\eta}_i)' (\boldsymbol{\Pi} \bar{\mathbf{s}}_t + \boldsymbol{\epsilon}_t) = \bar{\mathbf{s}}'_i \boldsymbol{\Theta} \bar{\mathbf{s}}_t + \boldsymbol{\theta}'_i \bar{\mathbf{s}}_t + \bar{\mathbf{s}}'_i \boldsymbol{\theta}_t + \boldsymbol{\eta}'_i \boldsymbol{\epsilon}_t,$$

where  $\boldsymbol{\Theta} = \boldsymbol{\Psi}' \boldsymbol{\Pi}$  is a  $p \times p$  matrix,  $\boldsymbol{\theta}_i = (\boldsymbol{\eta}'_i \boldsymbol{\Pi})'$  is a  $p \times 1$  vector, and  $\boldsymbol{\theta}_t = (\boldsymbol{\Psi}' \boldsymbol{\epsilon}_t)'$  is a  $p \times 1$  vector. Under the assumption that  $\boldsymbol{\eta}'_i \boldsymbol{\epsilon}_t$  and the right hand side variables are independent, we can augment the model and consistently estimate the effects of interest using an approach similar to the one presented below (see, e.g., Bai 2009, Abrevaya and Dahl 2008).

**Example 3.** The one-factor model case suggests that the coefficient  $\boldsymbol{\delta}(\tau)$  in equation 3.3 may be interpreted as a reduced form coefficient. From equation (3.4), we can write,

$$(3.14) \quad f_t = \boldsymbol{\Gamma}'_0 \bar{\mathbf{z}}_t(\tau) - \boldsymbol{\Gamma}'_0 \mathbf{C}_1 \bar{\mathbf{w}}_t - \boldsymbol{\Gamma}'_0 \mathbf{C}_2(\tau) \bar{\mathbf{x}}_t - \boldsymbol{\Gamma}'_0 \mathbf{C}_5 \bar{\boldsymbol{\lambda}},$$

where as before  $\boldsymbol{\Gamma}_0 = \tilde{\mathbf{C}}_3'(\tilde{\mathbf{C}}_3' \tilde{\mathbf{C}}_3)^{-1}$  and  $\tilde{\mathbf{C}}_3 = \bar{\mathbf{C}}_3 + \mathbf{C}_4 \bar{\boldsymbol{\lambda}}'$ . Multiplying (3.14) by the one-dimensional loading  $\lambda_i$ , we have that,

$$(3.15) \quad \lambda_i f_t = \lambda_i \boldsymbol{\Gamma}'_0 \bar{\mathbf{z}}_t(\tau) - \lambda_i \boldsymbol{\Gamma}'_0 \mathbf{C}_1 \bar{\mathbf{w}}_t - \lambda_i \boldsymbol{\Gamma}'_0 \mathbf{C}_2(\tau) \bar{\mathbf{x}}_t - \lambda_i \boldsymbol{\Gamma}'_0 \mathbf{C}_5 \bar{\boldsymbol{\lambda}}$$

$$(3.16) \quad = \boldsymbol{\delta}'_0 \bar{\mathbf{z}}_t(\tau) + \boldsymbol{\delta}'_1 \bar{\mathbf{w}}_t + \boldsymbol{\delta}'_2(\tau)' \bar{\mathbf{x}}_t + \boldsymbol{\delta}'_5 \bar{\boldsymbol{\lambda}},$$

where  $\boldsymbol{\delta}_0 = \boldsymbol{\Gamma}_0 \lambda_i$ ,  $\boldsymbol{\delta}_1 = -\boldsymbol{\delta}_0 \mathbf{C}_1$ ,  $\boldsymbol{\delta}_2(\tau) = -\boldsymbol{\delta}_0 \mathbf{C}_2(\tau)$ , and  $\boldsymbol{\delta}_5 = -\boldsymbol{\delta}_0 \mathbf{C}_5$ . Note that the reduced form parameter  $\boldsymbol{\delta}_0$  is then,

$$\boldsymbol{\delta}_0 = \boldsymbol{\Gamma}_0 \lambda_i = \frac{1}{N \tilde{\mathbf{C}}_3' \tilde{\mathbf{C}}_3} \sum_{i=1}^N \begin{pmatrix} \boldsymbol{\alpha}' \Pi_{3i} \lambda_i + (\boldsymbol{\alpha}' \Pi_4 + 1) \lambda_i^2 \\ \Pi_{3i} \lambda_i + \Pi_4 \lambda_i^2 \end{pmatrix} = \frac{1}{\tilde{\mathbf{C}}_3' \tilde{\mathbf{C}}_3} \begin{pmatrix} \boldsymbol{\alpha}' \bar{\lambda}_{\Pi_3} + (\boldsymbol{\alpha}' \Pi_4 + 1) \bar{\lambda}^2 \\ \bar{\lambda}_{\Pi_3} + \Pi_4 \bar{\lambda}^2 \end{pmatrix}.$$

where  $\bar{\lambda}_{\Pi_3} = N^{-1} \sum_i \Pi_{3i} \lambda_i$  and  $\bar{\lambda}^2 = N^{-1} \sum_i \lambda_i^2$ . Using equation (3.16) and letting  $\mathbf{f}(\tau) = (\bar{\mathbf{z}}_t(\tau)', \bar{\mathbf{w}}_t', \bar{\mathbf{x}}_t', \boldsymbol{\nu})'$  and  $\boldsymbol{\delta}(\tau) = (\boldsymbol{\delta}'_0, \boldsymbol{\delta}'_1, \boldsymbol{\delta}'_2(\tau)', \boldsymbol{\delta}'_5)'$ , we have that  $\lambda_i f_t$  equals  $f_t(\tau) \boldsymbol{\delta}(\tau)$ .

**Remark 2.** The conditional quantile function 3.1 allows for unobserved time-varying effects  $\mathbf{f}_t$ , and individual specific effects  $\boldsymbol{\lambda}_i$ , representing a more general version of a panel data quantile regression model. Consider first that the error terms  $u$  and  $v$  in model 2.1-2.2 are stochastically independent. By setting  $\mathbf{f}_t = r = 1$  for all  $t$ , we have the conditional quantile function  $Q_{Y_{it}}(\tau | \mathbf{d}_{it}, \mathbf{x}_{it}, \lambda_i) = \mathbf{d}'_{it} \boldsymbol{\alpha}(\tau) + \mathbf{x}'_{it} \boldsymbol{\beta}(\tau) + \lambda_i$  estimated in Koenker (2004) and Lamarche (2010). Moreover, if  $\boldsymbol{\lambda}_i = r = 1$  for all  $i$ , we have a conditional quantile function with time effects  $Q_{Y_{it}}(\tau | \mathbf{d}_{it}, \mathbf{x}_{it}, f_t) = \mathbf{d}'_{it} \boldsymbol{\alpha}(\tau) + \mathbf{x}'_{it} \boldsymbol{\beta}(\tau) + f_t$ . A simple variation arises when  $\mathbf{f}'_t \boldsymbol{\lambda}_i = \lambda_i + f_t$ , which can also be simply estimated by the method developed by Koenker. A model with endogenous covariates can be estimated by the method proposed in Harding and Lamarche (2009).

**Remark 3.** The recent literature on panel data quantile regression proposes to estimate a vector of  $N$  (nuisance) individual effects (see, e.g., Koenker 2004, Lamarche 2010). As recognized by Koenker (2004), the sparsity of the design plays a crucial role in the estimation procedure. Our approach is computationally simple, and we believe it is convenient for estimating large microeconomic panels. The procedure uses the R functions in `quantreg` (Koenker, 2010) and the implementation of the instrumental variable strategy follows closely the grid approach presented in Chernozhukov and Hansen (2006). We define a grid  $j = \{1, \dots, J\}$ . For a given  $j$ , we define a grid running ordinary quantile regression of  $y_{it} - \mathbf{d}'_{it} \boldsymbol{\alpha}_j$  on  $\mathbf{x}_{it}$ ,  $\hat{\mathbf{f}}_t(\tau)$ , and  $\hat{\boldsymbol{\Phi}}_{it}(\tau)$ , obtaining  $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}_j(\tau), \tau)$ ,  $\hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\alpha}}_j(\tau), \tau)$ ,  $\hat{\boldsymbol{\gamma}}(\hat{\boldsymbol{\alpha}}_j(\tau), \tau)$ . Then, we find  $\hat{\boldsymbol{\alpha}}(\tau) = \hat{\boldsymbol{\alpha}}_j^*$ , where  $\hat{\boldsymbol{\alpha}}_j^*$  minimizes  $\|\hat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha})\|_{\hat{\mathbf{A}}(\tau)}^2$ .

## 4. Additional Assumptions and Basic Inference

We now briefly state a series of results to facilitate the estimation of standard errors and evaluation of basic hypotheses. Consider the following assumptions:

**ASSUMPTION 6.** *The variables  $y_{it}$  are independent with conditional distribution  $G_{it}$ , and continuous densities  $g_{it}$  uniformly bounded away from 0 and  $\infty$ , with bounded derivatives  $g'_{it}$  everywhere.*

**ASSUMPTION 7.** *For all  $\tau \in \mathcal{T}$ ,  $(\boldsymbol{\alpha}(\tau), \boldsymbol{\beta}(\tau)) \in \text{int } \mathcal{A} \times \mathcal{B}$ , where  $\mathcal{A} \times \mathcal{B}$  is compact and convex.*

**ASSUMPTION 8.** *Let  $\boldsymbol{\pi} \equiv (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\delta}')'$  and  $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\delta}')'$ . Also,*

$$\boldsymbol{\Pi}(\boldsymbol{\pi}, \tau) \equiv \mathbb{E}[(\tau - 1)\{Y < \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{F}\boldsymbol{\delta}\}\boldsymbol{\Delta}(\tau)]$$

$$\boldsymbol{\Pi}(\boldsymbol{\theta}, \tau) \equiv \mathbb{E}[(\tau - 1)\{Y < \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\delta}\}\boldsymbol{\Delta}(\tau)]$$

where  $\boldsymbol{\Delta}(\tau) = [\mathbf{X}, \mathbf{W}, \mathbf{F}]'$ . The Jacobian matrices  $\partial\boldsymbol{\Pi}(\boldsymbol{\theta}, \tau)/\partial(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\delta}')$  and  $\partial\boldsymbol{\Pi}(\boldsymbol{\pi}, \tau)/\partial(\boldsymbol{\beta}', \boldsymbol{\delta}', \boldsymbol{\gamma}')$  have full rank and are continuous uniformly over  $\mathcal{A} \times \mathcal{B} \times \mathcal{L} \times \mathcal{G}$  and the image of  $\mathcal{A} \times \mathcal{B} \times \mathcal{L} \times \mathcal{G}$  under the mapping  $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}) \mapsto \boldsymbol{\Pi}(\boldsymbol{\theta}, \tau)$  is simply connected.

**ASSUMPTION 9.** *There exist limiting positive definite matrices  $\mathbf{S}(\tau)$  and  $\mathbf{J}(\tau)$  equal to,*

$$\mathbf{S}(\tau) = \lim_{\substack{T \rightarrow \infty \\ N \rightarrow \infty}} \frac{\tau(1-\tau)}{NT} \tilde{\mathbf{X}}' \mathbf{M}'_F \mathbf{M}_F \tilde{\mathbf{X}}$$

$$\mathbf{J}(\tau) = \lim_{\substack{T \rightarrow \infty \\ N \rightarrow \infty}} \frac{1}{NT} (\mathbf{K}', \mathbf{L}')$$

where  $\tilde{\mathbf{X}} = [\mathbf{W}, \mathbf{X}]$ ,  $\mathbf{M}_F = \mathbf{I} - \mathbf{P}_F$ ,  $\mathbf{P}_F = \mathbf{F}(\mathbf{F}'\boldsymbol{\Phi}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Phi}$ ,  $\boldsymbol{\Phi} = \text{diag}(g_{it}(\xi_{it}(\tau)))$ ,  $\mathbf{K} = [\mathbf{J}'_\alpha \mathbf{H} \mathbf{J}_\alpha]^{-1} \mathbf{J}'_\alpha \mathbf{H}$ ,  $\mathbf{J}_\alpha = \lim_{N, T \rightarrow \infty} \tilde{\mathbf{X}}' \mathbf{M}'_F \boldsymbol{\Phi} \mathbf{M}_F \mathbf{D}$ ,  $\mathbf{H} = \bar{\mathbf{J}}'_\gamma \mathbf{A} \bar{\mathbf{J}}_\gamma$ ,  $\mathbf{A}$  is a positive definite matrix,  $\mathbf{L} = \bar{\mathbf{J}}_\alpha \mathbf{M}$ ,  $\mathbf{M} = \mathbf{I} - \mathbf{J}_\alpha \mathbf{K}$ , and  $(\bar{\mathbf{J}}_\alpha, \bar{\mathbf{J}}_\gamma)$  are partitions of the inverse of  $\mathbf{J}_\vartheta = \lim_{N, T \rightarrow \infty} \tilde{\mathbf{X}}' \mathbf{M}'_F \boldsymbol{\Phi} \mathbf{M}_F \tilde{\mathbf{X}}$ .

**ASSUMPTION 10.**  $\max \|\mathbf{x}_{it}\|/\sqrt{NT} \rightarrow 0$ ,  $\max \|\mathbf{f}_t\|/\sqrt{N} \rightarrow 0$ , and  $\max \|\mathbf{w}_{it}\|/\sqrt{NT} \rightarrow 0$ .

The previous conditions are standard in the literature. 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 6 ensures a well-defined asymptotic behavior of the quantile regression estimator. Condition 4 implies the standard conditions on the instruments in the exactly identified case. This case can be easily relaxed but we impose it by convenience. The independence on the  $y_{it}$ 's is assumed in Koenker (2004) and conditions 7 and 8 are assumed in Chernozhukov and Hansen (2006). In condition 9, the existence of the limiting form of the positive definite matrices is used to invoke the

Lindeberg-Feller Central Limit Theorem. The asymptotic covariance matrices have a representation similar to the matrices found in Chernozhukov and Hansen (2008, proposition 2) and Galvao (2009, theorem 2). Condition 10 is important both for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function.

**PROPOSITION 2.** *Under regularity conditions A1-A10, the quantile regression estimator for a model with interactive effects,  $(\hat{\alpha}(\tau)', \hat{\beta}(\tau)')$ , is consistent and asymptotically normally distributed with mean  $(\alpha(\tau)', \beta(\tau)')$  and covariance matrix  $\mathbf{J}'(\tau)\mathbf{S}(\tau)\mathbf{J}(\tau)$ .*

Estimation and basic inference requires us to approximate  $\mathbf{F}$  and estimate the conditional density  $g$ . The matrix of factors  $\mathbf{F}$  of dimension  $NT \times r$  is replaced by a matrix  $\hat{\mathbf{F}}$  of dimension  $NT \times 1 + k_1 + k_2 + m$ . This matrix includes cross-sectional averages as in Proposition 1, although it is possible to construct  $\hat{\mathbf{F}}(\tau)$  that incorporate quantile-specific variables. The density  $g$  can be estimated by employing standard quantile regression techniques. To estimate the density, we adopt a kernel estimator using residuals of the form  $\hat{u}_{it}(\tau) = y_{it} - \mathbf{d}'_{it}\hat{\alpha}(\tau) - \mathbf{x}'_{it}\hat{\beta}(\tau) - \hat{\mathbf{f}}'_t\hat{\delta}(\tau)$  and a properly chosen bandwidth  $h$  (see Koenker 2005 for details). Therefore, the matrix  $\mathbf{S}(\tau)$  can be estimated by  $\hat{\mathbf{S}}(\tau) = \tau(1 - \tau)(NT)^{-1}\tilde{\mathbf{X}}'\hat{\mathbf{M}}'_{\hat{\mathbf{F}}}\hat{\mathbf{M}}_{\hat{\mathbf{F}}}\tilde{\mathbf{X}}$ , where  $\tilde{\mathbf{X}} = [\mathbf{W}, \mathbf{X}]$  and  $\hat{\mathbf{M}}_{\hat{\mathbf{F}}} = \mathbf{I} - \hat{\mathbf{F}}(\hat{\mathbf{F}}'\hat{\Phi}\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}'\hat{\Phi}$ . The matrix  $\mathbf{J}(\tau)$  can be similarly estimated.

Inference for panel data can be considered by accommodating standard quantile regression tests. Wald-type statistics (see, e.g., Koenker and Bassett 1982) can be considered for basic general linear hypothesis on a vector  $\boldsymbol{\theta}$ . More general hypothesis including evaluating the vector over a range of quantiles could be accommodated by considering the Kolmogorov-Smirnov statistics (see, e.g., Koenker and Xiao 2002, and Koenker 2005). We consider a general linear hypothesis on the vector  $\boldsymbol{\theta}$  of the form  $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ , where  $\mathbf{R}$  is a matrix that depends on the type of restrictions imposed. This framework allows us to evaluate several linear hypotheses, including the significance of differences across coefficient estimates of a vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1(\tau), \dots, \boldsymbol{\theta}_p(\tau))'$ . The test statistics is defined as,

$$(4.1) \quad \mathcal{T}_{NT} = NT(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r})' \left[ \mathbf{R}\hat{\mathbf{V}}^{-1}\mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}),$$

is asymptotically distributed as  $\chi_q^2$  under  $H_0$ , where  $q$  is the rank of the matrix  $\mathbf{R}$  and  $\hat{\mathbf{V}}$  is the estimated covariance matrix,  $\hat{\mathbf{J}}'(\tau)\hat{\mathbf{S}}(\tau)\hat{\mathbf{J}}(\tau)$ .

## 5. Monte Carlo

In this section, we report the results of several simulation experiments designed to evaluate the performance of the method in finite samples. We compare the small sample behavior of the estimator proposed in this paper within a class of fixed effects methods. First, we will briefly investigate the bias and root mean squared error (RMSE) of the estimator in models where the endogenous variable is correlated with the unobserved factors. Second, we investigate the performance of the method when the endogenous variable is correlated with both loadings and factors. Lastly, we provide evidence on the performance of the method when the endogenous variable is correlated with the loadings, factors and the error term.

We generate the dependent variable considering a design similar to Bai (2009) and Pesaran (2006):

$$(5.1) \quad y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + (1 + h d_{it}) u_{it}$$

$$(5.2) \quad d_{it} = \pi_0 + \pi_1 w_{it} + \pi_2 x_{it} + \pi_3 f_{1t} + \pi_3 f_{2t} + \pi_4 \lambda_{1i} f_{1t} + \pi_4 \lambda_{2i} f_{2t} + v_{it}$$

$$(5.3) \quad w_{it} = \theta_0 \mu_t + \theta_1 \epsilon_{it},$$

$$(5.4) \quad f_{jt} = \rho_f f_{jt-1} + \eta_{jt}$$

$$(5.5) \quad \eta_{jt} = \rho_\eta \eta_{jt-1} + e_{jt}$$

for  $j = \{1, 2\}, \dots, t = -49, \dots, 0, \dots, T$  in the last two equations. The error terms are  $(u_{it}, v_{it})' \sim \mathcal{N}(0, \mathbf{\Omega})$  and  $x_{it}, \mu_t, \epsilon_{it}, e_{it}$  are Gaussian independent random variables. The loading  $\lambda_{i1} \sim \mathcal{N}(1, 0.2)$  and  $\lambda_{i2}$  is distributed either as  $\mathcal{N}(1, 0.2)$  or  $t$ -student distribution with two degrees of freedom. The parameters are assumed to be:  $\beta_0 = \pi_0 = 0, \beta_1 = \beta_2 = \pi_1 = \pi_2 = \pi_3 = \theta_0 = \theta_1 = 1, \rho_f = 0.90, \rho_\eta = 0.25$ , and  $\Omega_{11} = \Omega_{22} = 1$ .

We consider three designs for the location shift model  $h = 0$ :

**Design 1:** The endogenous variable  $d$  is not correlated with the  $\lambda$ 's, and the variables  $u$  and  $v$  are independent Gaussian variables. Although  $d$  is not correlated with the individual effects and the error term, it is correlated with the  $F$ 's. We assume  $\pi_4 = 0$  and  $\Omega_{12} = \Omega_{21} = 0$ .

**Design 2:** The variable  $d$  is correlated with  $F$ 's and  $\lambda$ 's, and the error terms in equations 5.1 and 5.2 are not correlated. We assume  $\pi_4 = 0.1$  and  $\Omega_{12} = \Omega_{21} = 0$ .

**Design 3:** The error terms in equations 5.1 and 5.2 are now correlated, assuming that  $\Omega_{12} = \Omega_{21} = 0.5$ . The variable  $d$  is also correlated with the  $F$ 's and  $\lambda$ 's as in the experiment carried out in Design 2.

Moreover, we expand the analysis considering these designs for the case of location-scale shift models. In all the variants of the experiments, we assume that  $h = 0.1$  in equation 5.1.

Statistic	Sample Size		Panel Data Methods							
			Least Squares				Quantile Regression			
	N	T	OLS	LSFE	LSIVFE	LSIE	QR	QRFE	QRIVFE	QRIE
			Gaussian distribution							
Bias	100	4	0.2652	0.3013	-0.2325	-0.0012	0.2695	0.3038	-0.2319	0.0092
RMSE	100	4	0.3297	0.3671	2.1426	0.0901	0.3352	0.3700	1.6578	0.1706
Bias	250	4	0.2751	0.3112	-0.0879	0.0001	0.2766	0.3141	-0.1863	-0.0010
RMSE	250	4	0.3314	0.3717	5.2631	0.0564	0.3345	0.3751	4.6359	0.1048
Bias	100	8	0.4133	0.4317	-0.0937	-0.0012	0.4106	0.4230	-0.1291	-0.0138
RMSE	100	8	0.4514	0.4700	0.6048	0.0634	0.4498	0.4609	0.8426	0.1404
Bias	250	8	0.3898	0.4100	-0.1104	0.0006	0.3864	0.4002	-0.1341	-0.0004
RMSE	250	8	0.4284	0.4480	1.0372	0.0403	0.4251	0.4382	1.1526	0.0839
			$t_2$ distribution							
Bias	100	4	0.1543	0.1789	-0.0573	0.0160	0.1542	0.1783	-0.0664	0.0098
RMSE	100	4	0.2629	0.2665	0.3911	0.3514	0.2337	0.2567	0.4506	0.2324
Bias	250	4	0.1299	0.1497	-0.2558	-0.0036	0.1343	0.1525	-0.2122	0.0021
RMSE	250	4	0.2181	0.2390	2.7363	0.2393	0.2033	0.2300	1.9292	0.1626
Bias	100	8	0.1929	0.2118	-0.0062	-0.0386	0.2009	0.2154	-0.0616	-0.0075
RMSE	100	8	0.2780	0.2849	0.8160	0.2481	0.2703	0.2753	0.6111	0.2142
Bias	250	8	0.2099	0.2206	-0.1737	0.0074	0.2126	0.2244	-0.2297	0.0057
RMSE	250	8	0.2765	0.2886	0.7196	0.1426	0.2699	0.2792	0.8133	0.1110

TABLE 5.1. *Small sample performance of a class of panel data estimators. This table considers the Monte Carlo design 1. The table includes bias and root mean square error (RMSE) for the slope parameter. The evidence is based on 400 randomly generated samples.*

Tables 5.1-5.3 present the bias and root mean square error (RMSE) of the simulation experiments. While the upper panels of the tables present results for the case that  $\lambda_1$  and  $\lambda_2$  are distributed as Gaussian random variables ( $\mathcal{N}(1, 0.2)$ ), the lower panels of the tables present results for the case of  $\lambda_2$  distributed as  $t$ -student distribution with two degrees of freedom ( $t_2$ ). We present results based on  $N = \{100, 250\}$  and  $T = \{4, 8\}$ . It is important to note that even though the asymptotic results were derived under the assumption of large  $N$  and large  $T$ , the finite sample performance of the proposed estimators is excellent even for a very small number of time periods. The tables show

Statistic	Sample Size		Panel Data Methods							
			Least Squares				Quantile Regression			
			OLS	LSFE	LSIVFE	LSIE	QR	QRFE	QRIVFE	QRIE
N T		Gaussian distribution								
Bias	100	4	0.3302	0.3148	-0.2153	-0.0018	0.3311	0.3169	-0.2621	0.0085
RMSE	100	4	0.3764	0.3726	2.1016	0.0901	0.3783	0.3746	2.1627	0.1721
Bias	250	4	0.3442	0.3245	0.0692	-0.0001	0.3417	0.3264	-0.0582	-0.0024
RMSE	250	4	0.3876	0.3787	2.8627	0.0564	0.3866	0.3813	2.7423	0.1088
Bias	100	8	0.4530	0.4391	-0.1640	-0.0015	0.4480	0.4304	-0.1850	-0.0136
RMSE	100	8	0.4820	0.4714	1.3625	0.0633	0.4781	0.4624	1.3522	0.1382
Bias	250	8	0.4370	0.4213	-0.4092	0.0005	0.4300	0.4115	-0.4482	0.0016
RMSE	250	8	0.4648	0.4528	4.8337	0.0402	0.4583	0.4433	5.4814	0.0853
N T		$t_2$ distribution								
Bias	100	4	0.9599	0.3774	-0.0734	0.0039	0.4254	0.2372	-0.0784	0.0141
RMSE	100	4	1.5011	0.6422	0.4356	0.3548	0.5365	0.2937	0.5127	0.2435
Bias	250	4	1.1461	0.3508	-0.0676	-0.0092	0.4620	0.2119	-0.0897	0.0092
RMSE	250	4	1.7067	0.5390	1.0215	0.2409	0.6094	0.2702	0.8594	0.1622
Bias	100	8	0.9047	0.4055	-0.0502	-0.0450	0.4398	0.2792	-0.0949	-0.0064
RMSE	100	8	1.2719	0.4856	0.5867	0.2622	0.5145	0.3215	0.5524	0.2101
Bias	250	8	0.9443	0.4435	-0.2747	0.0054	0.4421	0.2925	-0.3328	0.0029
RMSE	250	8	1.3166	0.5388	1.5394	0.1426	0.5153	0.3272	1.4542	0.1198

TABLE 5.2. *Small sample performance of a class of panel data estimators. This table considers the Monte Carlo design 2. The table includes bias and root mean square error (RMSE) for the slope parameter. The evidence is based on 400 randomly generated samples.*

results obtained from: ordinary least squares (OLS), the least squares version of the fixed effects estimator (LSFE), the classical instrumental variable estimation for a model with fixed effects (LSIVFE), the estimator introduced in proposition 1 (LSIE) which we consider to be a version of the CCE estimator of Pesaran (2006), the classical quantile regression estimator (QR), the fixed effects version of the estimator introduced by Koenker (2004) labelled QRFE, the instrumental variable method with fixed effects (QRIVFE) introduced by Harding and Lamarche (2009), and the quantile regression estimator for a model with interactive effects (QRIE).

Statistic	Sample Size		Panel Data Methods							
			Least Squares				Quantile Regression			
			OLS	LSFE	LSIVFE	LSIE	QR	QRFE	QRIVFE	QRIE
N T		Gaussian distribution								
Bias	100	4	0.4708	0.4413	-0.2048	0.0039	0.4710	0.4410	-0.2639	0.0149
RMSE	100	4	0.4948	0.4711	2.0380	0.0870	0.4962	0.4711	2.1051	0.1806
Bias	250	4	0.4757	0.4414	0.0740	0.0021	0.4714	0.4397	-0.0582	0.0022
RMSE	250	4	0.5007	0.4726	2.8873	0.0591	0.4971	0.4712	2.7774	0.1022
Bias	100	8	0.5521	0.5318	-0.1604	0.0059	0.5459	0.5229	-0.1736	0.0092
RMSE	100	8	0.5693	0.5509	1.3783	0.0638	0.5636	0.5422	1.3123	0.1331
Bias	250	8	0.5402	0.5176	-0.4246	0.0008	0.5344	0.5069	-0.4783	0.0061
RMSE	250	8	0.5561	0.5355	5.1030	0.0396	0.5508	0.5252	5.8590	0.0889
N T		$t_2$ distribution								
Bias	100	4	1.0831	0.4939	-0.0753	0.0056	0.5513	0.3502	-0.0906	0.0036
RMSE	100	4	1.5733	0.7102	0.4294	0.3546	0.6415	0.3821	0.5747	0.2676
Bias	250	4	1.2690	0.4710	-0.0742	-0.0107	0.5937	0.3330	-0.1188	0.0041
RMSE	250	4	1.7823	0.6206	1.1007	0.2410	0.7166	0.3678	0.9933	0.1731
Bias	100	8	0.9985	0.4990	-0.0531	-0.0472	0.5387	0.3785	-0.0856	-0.0067
RMSE	100	8	1.3388	0.5635	0.5714	0.2657	0.6002	0.4061	0.5913	0.1955
Bias	250	8	1.0402	0.5384	-0.2735	0.0037	0.5437	0.3912	-0.3240	0.0077
RMSE	250	8	1.3834	0.6171	1.5494	0.1451	0.6019	0.4130	1.4801	0.1247

TABLE 5.3. *Small sample performance of a class of panel data estimators. This table considers the Monte Carlo design 3. The table includes bias and root mean square error (RMSE) for the slope parameter. The evidence is based on 400 randomly generated samples.*

Tables 5.1 and 5.2 provide evidence of the biases present in the application of traditional methods in the presence of interactive effects under alternative assumptions on how the unobserved time-varying interactive effects are correlated with other right hand side variables. While it may not be surprising to find out that pooled methods such as OLS or QR are biased and have poor MSE properties, it is surprising that the application of standard FE estimation to a situation with interactive effects produces large biases. This is true for both mean regression and quantile regression. In particular notice how the differencing procedure applied to mean regression analysis with fixed effects fails to remove the unobserved heterogeneity and induces additional correlations which lead to a large bias and MSE. Applying an instrumental variables technique, in addition



Statistic	Sample Size		0.25 Quantile				0.75 Quantile			
	QR	QRFE	QRIVFE	QRIE	QR	QRFE	QRIVFE	QRIE		
Monte Carlo Design 1 - Gaussian distribution										
	N	T								
Bias	100	4	0.2934	0.3217	-0.1348	0.0468	0.2311	0.2732	-0.3608	-0.0354
RMSE	100	4	0.3605	0.3891	1.1409	0.1905	0.3101	0.3459	2.8677	0.1845
Bias	250	4	0.3075	0.3336	-0.0708	0.0342	0.2383	0.2778	-0.1999	-0.0394
RMSE	250	4	0.3666	0.3963	4.3282	0.1245	0.3054	0.3428	4.1701	0.1338
Bias	100	8	0.4478	0.4537	-0.1360	0.0370	0.3729	0.3798	-0.1181	-0.0309
RMSE	100	8	0.4897	0.4940	0.8972	0.1462	0.4173	0.4200	0.6015	0.1536
Bias	250	8	0.4192	0.4205	-0.1289	0.0305	0.3537	0.3697	-0.1800	-0.0297
RMSE	250	8	0.4616	0.4589	0.8784	0.0977	0.3967	0.4097	1.0934	0.1077
Monte Carlo Design 2 - Gaussian distribution										
	N	T								
Bias	100	4	0.3590	0.3337	-0.1508	0.0496	0.3008	0.2867	-0.2563	-0.0395
RMSE	100	4	0.4101	0.3924	1.4565	0.1922	0.3551	0.3507	1.9657	0.1983
Bias	250	4	0.3742	0.3459	0.0244	0.0321	0.3058	0.2921	-0.0010	-0.0383
RMSE	250	4	0.4208	0.4021	2.0425	0.1302	0.3549	0.3505	3.6376	0.1389
Bias	100	8	0.4858	0.4609	-0.1827	0.0404	0.4099	0.3894	-0.1420	-0.0372
RMSE	100	8	0.5182	0.4949	1.3417	0.1548	0.4444	0.4241	0.7657	0.1606
Bias	250	8	0.4633	0.4313	-0.3888	0.0316	0.3987	0.3817	-0.3184	-0.0306
RMSE	250	8	0.4955	0.4638	4.7955	0.0997	0.4299	0.4150	1.6772	0.1077
Monte Carlo Design 3 - Gaussian distribution										
	N	T								
Bias	100	4	0.5018	0.4641	-0.1604	0.0247	0.4267	0.4002	-0.2308	-0.0141
RMSE	100	4	0.5316	0.5003	1.6530	0.1910	0.4610	0.4377	1.7873	0.1922
Bias	250	4	0.5136	0.4689	0.0058	0.0358	0.4279	0.4008	-0.0024	-0.0345
RMSE	250	4	0.5423	0.5041	2.1258	0.1278	0.4622	0.4361	3.6637	0.1452
Bias	100	8	0.5896	0.5582	-0.1956	0.0218	0.4966	0.4725	-0.1388	-0.0068
RMSE	100	8	0.6094	0.5801	1.3326	0.1587	0.5215	0.4950	0.7900	0.1539
Bias	250	8	0.5728	0.5314	-0.4110	0.0185	0.4925	0.4667	-0.3028	-0.0138
RMSE	250	8	0.5916	0.5523	5.0595	0.0979	0.5139	0.4878	1.5624	0.1089

TABLE 5.4. *Small sample performance of quantile regression estimators in the location-scale shift model. This table considers the Monte Carlo designs 1-3. The table includes bias and root mean square error (RMSE) for the slope parameter.*

to the fixed effects procedure, partially ameliorates the previously documented biases, but not completely. This approach to addressing the interactive nature of the unobserved heterogeneity comes at a large cost in terms of estimator efficiency. The proposed quantile regression estimator has almost zero bias and a low MSE. Additionally we notice that the performance of the least squares LSIE estimator deteriorates in the presence of outliers, i.e. if the loading  $\lambda_2$  follows a  $t_2$  distribution, while the quantile estimator QRIE continues to perform well with only a minor increase in MSE.

Table 5.3 confirms that the above interpretation of the relative performance of the different estimators persists in the more difficult case which features both endogeneity and interactive effects. The method proposed in this paper QRIE continues to perform very well in this case having very low bias and MSE even when the number of time periods under observation is small. Moreover, by comparing the least squares simulations with the quantile regression simulations we can see that a quantile regression based approach such as QRIE delivers substantially better MSE properties.

Table 5.4 presents the bias and root mean square error in the Gaussian case when  $\beta_1$  in equation 5.1 represents a location-scale shift. We present results at the 0.25 and 0.75 quantiles. We find that the bias of the estimator in the simulations is very low, ranging from 0.7 percent to 5.0 percent in absolute value. The performance of the QRIE estimator continues to be satisfactory, offering in general the smallest bias and best MSE in the class of fixed effects estimators presented in the table.

Lastly, we expand the previous designs in Table 5.5 by replacing equation 5.1 by:

$$(5.6) \quad y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 x_t + c_1(\lambda_{1i} f_{1t} + \lambda_{2i} f_{2t}) + c_2 \lambda_{3i} + (1 + h d_{it}) u_{it},$$

with  $\lambda_{i3}$  distributed as *iid* normal random variable with mean 0 and variance 1. We consider three cases of the location-scale shift model with  $h = 0.1$ :

**Design 4:** The dependent variable is generated from the classical model with individual effects. We assume that  $c_1 = 0$  and  $c_2 = 1$ .

**Design 5:** We introduce a location-scale shift model with individual and interactive effects. We assume that  $c_1 = c_2 = 1$ .

**Design 6:** The dependent variable is generated from a location-scale shift model with an endogenous variable,  $d$ . It is assumed that  $\Omega_{21} = \Omega_{12} = 0.5$ . The model also includes individual and interactive effects, because  $c_1 = c_2 = 1$ .

Statistic	Sample		0.25 Quantile				0.75 Quantile			
	N	T	QRFE	QRIVFE	QRIE	QRIIE	QRFE	QRIVFE	QRIE	QRIIE
Design 4: Model with individual effects										
Bias	100	4	0.0230	0.0815	0.0529	0.0358	-0.0134	-0.0824	-0.0221	-0.0262
RMSE	100	4	0.0481	1.2867	0.1545	0.1440	0.0493	1.0241	0.1445	0.1482
Bias	100	6	0.0183	0.0027	0.0266	0.0283	-0.0133	-0.0185	-0.0178	-0.0365
RMSE	100	6	0.0355	0.2244	0.1265	0.1105	0.0334	0.2244	0.1537	0.1143
Bias	100	12	0.0076	-0.0464	0.0307	0.0221	-0.0086	-0.1739	-0.0283	-0.0257
RMSE	100	12	0.0195	1.1278	0.0994	0.0955	0.0210	2.2456	0.1120	0.0980
Bias	100	20	0.0065	0.0564	0.0312	0.0212	-0.0067	-0.0543	-0.0334	-0.0254
RMSE	100	20	0.0134	0.3401	0.0986	0.0763	0.0134	0.2640	0.0911	0.0750

Design 5: Model with individual effects and interactive effects

Bias	100	4	0.3720	0.0419	0.0537	0.0467	0.3051	-0.3001	-0.0006	-0.0196
RMSE	100	4	0.4305	2.7468	0.2222	0.1935	0.3692	3.3622	0.2120	0.2002
Bias	100	6	0.3826	-0.1634	0.0338	0.0355	0.3227	-0.1244	-0.0274	-0.0443
RMSE	100	6	0.4258	0.9781	0.1990	0.1321	0.3707	0.9931	0.1999	0.1532
Bias	100	12	0.5082	-1.6556	0.0359	0.0271	0.4580	-2.7600	-0.0288	-0.0296
RMSE	100	12	0.5318	24.9890	0.1342	0.1167	0.4836	40.8708	0.1577	0.1303
Bias	100	20	0.5972	-0.0713	0.0360	0.0315	0.5328	-0.2495	-0.0399	-0.0335
RMSE	100	20	0.6133	0.8050	0.1125	0.1010	0.5490	1.0512	0.1147	0.1078

Design 6: Model with endogenous covariates and individual and interactive effects

Bias	100	4	0.4861	-0.0339	0.0253	0.0320	0.4075	-0.2240	-0.0130	-0.0049
RMSE	100	4	0.5228	2.3460	0.2099	0.2024	0.4484	4.1126	0.2124	0.2087
Bias	100	6	0.4918	-0.1659	0.0218	0.0161	0.4253	-0.1313	-0.0217	-0.0311
RMSE	100	6	0.5194	0.9095	0.1877	0.1394	0.4559	0.9775	0.2126	0.1554
Bias	100	12	0.6027	-1.7410	0.0129	0.0158	0.5263	-1.6913	-0.0106	-0.0216
RMSE	100	12	0.6174	26.3003	0.1437	0.1234	0.5434	24.0678	0.1625	0.1286
Bias	100	20	0.6655	-0.0660	0.0220	0.0194	0.5843	-0.2332	-0.0374	-0.0212
RMSE	100	20	0.6771	0.7962	0.1119	0.0974	0.5956	1.0021	0.1159	0.1088

TABLE 5.5. *Small sample performance of quantile regression estimators in the location-scale shift model. This table considers the Monte Carlo designs 4-6. The table includes bias and root mean square error (RMSE) for the slope parameter.*

As before, Table 5.5 shows results obtained from: the quantile regression fixed effects estimator (QRFE), the instrumental variable method with fixed effects (QRIVFE), and the quantile regression estimator for a model with interactive effects (QRIE). We also show results obtained from an extension of the QRIE estimator. The estimator, which is labelled as QRIIE, includes individual effects and interactive effects.

Table 5.5 presents the bias and root mean square error in panels of sample size  $N = 100$  and  $T = \{4, 6, 12, 20\}$ . As expected, the performance of the QRFE estimator is satisfactory in the model with individual effects. We note, however, that the bias of the QRFE estimator tends to disappear as  $T$  increases. On the other hand, the QRFE has a poor MSE performance in all the variants of these models with interactive effects. When the model includes endogenous covariates, individual and interactive effects, the QRIIE estimator offers the best small sample performance in terms of bias and MSE.

Overall, the finite sample performances of the methods for models with interactive effects is very good in all the variants of the models considered in Tables 5.1-5.5. The performance of the least squares method offered in proposition 1 (LSIE) is satisfactory. It produces unbiased results, and it offers the best small sample performance under Gaussian conditions. The LSIE method, however, relies heavily on the distributional assumptions and is not informative for estimating distributional effects. When we relax the Gaussian conditions, the quantile regression version of the interactive effects estimator (QRIE) has smaller RMSE than the least squares version of the estimator (LSIE). The QRIE estimator is unbiased in all the variants of the models and it offers the best performance in the class of fixed effects quantile regression estimators.

## 6. Educational outcomes and heterogeneous class size effects

In this section, we consider data from a random assignment of college students to different classes, to study how class size and socioeconomic class composition affect educational attainment using data from De Giorgi, Pellizzari and Woolston (2009).

We apply our quantile regression interactive effects method to a structural equation model of students' educational achievement where teachers' and students' unobserved heterogeneity are allowed to interact. We will then compare the policy implications of changes in the number of students in a class and changes in the class composition on educational achievement, obtained from a series of instrumental variable and panel data methods. We find that our method suggests different policies relative to other existing estimation approaches. The instrumental variable estimator suggests

that a reduction in class size has an insignificant effect at the tails of the conditional distribution. We find, however, that while for the worst students, smaller classes improve performance, for the strongest students, smaller classes reduce performance. This is consistent with current research in educational psychology which suggests that weaker students benefit from the increased attention they receive in smaller classes, while strong students flourish more in competitive environments and benefit from the numerous interactions that larger classes provide.

Lastly, we investigate an educational policy designed to prevent failure among low performers. The findings suggest that a reduction of 10% in the size of a class, could improve educational achievement from the lower 10 percentile of the conditional achievement distribution to the next 20 percentile.

## 6.1. Background

In the last half a century, understanding what drives students' academic performance has been a major focus in the economics of education. The analysis of class size reduction on educational attainment continues to be one of the controversial topics in the social sciences ever since the Coleman Report (1966). A number of more recent studies have focused on class size and peer effects (e.g., Krueger 1999, Hoxby 2000, Hanushek et al. 2003). The standard methodologies employed in the empirical analysis have been Ordinary Least Squares (OLS), Instrumental Variable (IV), and regression discontinuity design (RD) methods (see, e.g., Angrist and Pischke 1999, Hoxby 2000, Duflo, Dupas and Kremer 2008). Class size remains one of the main factors to have been analyzed in the educational production function, although the conclusions remain mixed. While a number of studies find that a reduction in class size is associated with positive effects on student performance, other studies suggest that the effects are negligible and not statistically significant. Most studies were carried out on primary school students yet the number of students enrolled in various degree-granting institutions has increased by over 26% in the U.S. over the past 10 years. In this paper we take advantage of a unique educational design at a university in Italy. Mirroring the U.S., enrollment in higher education has increased sharply throughout Europe, leading to larger class sizes.

The literature to date offers a large number of studies on the effect of class size on educational achievement of the average student, but few studies investigate its distributional effect. One of the few exceptions are Levin (2001), Ma and Koenker (2006), and more recently, Bandiera, Larcinese, and Rasul (2010). Levin (2001) addresses the potential endogeneity in the class size variable

with a two-stage quantile regression approach and a large number of observable characteristics. The study uses PRIMA data, a longitudinal survey containing information on Dutch pupils who were enrolled in grades 2, 4, 6 and 8 in the 1994/1995 school year. Ma and Koenker (2006) estimate a structural model using the PRIMA data. They find that for the weak students, larger classes improve achievement in languages and smaller classes improve achievement in mathematics. Finally, Bandiera, Larcinese, and Rasul (2010) employ standard quantile regressions on a sample of university students in the UK. They find that a reduction of class size benefits high-performers more than the low-performers.

A central concern in the estimation of the distributional effects of class size and class composition is unobserved heterogeneity. Although the literature recently addresses the possibility that class size may not have the same effect for weak and strong students, it remains possible that student latent heterogeneity and teachers' latent ability factors create biases. It may be also possible that teachers' quality affects performance *only if* the student is motivated and receptive to instruction. In the next sections, we estimate the distributional effects of potential college policies designed to affect educational achievement, considering that student's motivation and teacher's quality enter multiplicatively in the production function.

## 6.2. Data

We employ data from the administrative records of the Economics, Management, and Finance programs at Bocconi University. The university is an established higher education institution located in Milan, Italy. The administration extracted a comprehensive data set of college students including information on students' characteristics and outcomes. The data set includes information on course grades, background demographic and socioeconomic characteristics such as gender, family income, and pre-enrollment test scores. Additionally, the data set includes information on enrollment year, academic program, number of exams by academic year, official enrollment, official proportion of female students in each class, and official proportion of high income students in each class. We restrict our attention to students who matriculated in the 1999-2000 academic year taking non-elective classes in the first three years of the program.

Table 6.1 reports briefly the summary statistics of the variables used in this study. The average grade is 25.602, which can be associated to a *B+* in the US grading system (De Giorgi, Pellizzari and Woolston, 2009). The average class includes 134 students, ranging from 90 to 158 students. The average number of assigned students to these classes is 133, and ranges from 89 to 160 students.

Variable	Mean	Std. Dev.	Minimum	Maximum
Economics, Management, and Finance programs				
Grade	25.602	2.400	18.857	30.500
Class size	134.395	15.759	90	158
Enrollment	132.534	15.529	89	160
% female in class (actual)	0.360	0.066	0.209	0.511
% female in class (enrollment)	0.364	0.071	0.205	0.507
% high income in class (actual)	0.232	0.040	0.130	0.316
% high income in class (enrollment)	0.248	0.039	0.165	0.344
Female indicator	0.387	0.487	0.000	1.000
High income indicator	0.218	0.413	0.000	1.000
Entry test	73.272	13.783	22.620	108.350
Management, and Finance programs				
Grade	25.601	2.391	18.857	30.500
Class size	137.953	10.256	124	158
Enrollment	136.031	10.157	119	160
% female in class (actual)	0.358	0.068	0.209	0.511
% female in class (enrollment)	0.361	0.073	0.205	0.507
% high income in class (actual)	0.235	0.037	0.181	0.316
% high income in class (enrollment)	0.250	0.038	0.185	0.344
Female indicator	0.386	0.487	0.000	1.000
High income indicator	0.218	0.413	0.000	1.000
Entry test	73.002	13.655	22.620	108.350

TABLE 6.1. *Descriptive statistics. Our data is based on administrative records of three academic programs at Bocconi University. The administration assigned students to classes using lotteries. The variables enrollment, % female in class, and % high income in class will be used as instruments.*

The table indicates a small degree of self-selection into classes post-randomization which explains the difference between actual class enrollment and allocation. The table also indicates that the largest classes are in Management, and Finance. When we restrict our attention to a sample of students in these programs, the average class consists of 137 students, ranging from 124 to 158 students.

### 6.3. Model specification

We estimate a structural equation model of the effect of class size and socioeconomic class composition on educational attainment. Our basic strategy is to account for the unobserved components that are most likely to contaminate class size and class heterogeneity influences on educational

achievement. We estimate the following model:

$$(6.1) \quad y_{ict} = \mathbf{d}'_{ct} \boldsymbol{\alpha} + \mathbf{x}'_{ict} \boldsymbol{\beta} + \mathbf{z}'_t \boldsymbol{\gamma} + \mathbf{f}'_{ct} \boldsymbol{\lambda}_i + u_{ict}$$

$$(6.2) \quad \mathbf{d}_{ct} = \mathbf{w}'_{ct} \boldsymbol{\pi}_1 + \mathbf{x}'_i \boldsymbol{\pi}_2 + \mathbf{f}'_{ct} \boldsymbol{\lambda}_i + \mathbf{v}_{ict},$$

where  $y$  is the average grade of student  $i$  in a class  $c$  at year  $t$ , and  $\mathbf{d}$  is a vector of potentially endogenous variables that includes class size, and measures of actual dispersion of gender and income in each class. The vector  $\mathbf{x}$  includes indicators for gender, whether or not the student is a high income student, and a cognitive test score corresponding to a test the student took as part of the admission process. We include additional control variables such as the standard deviation of the logarithm of the test score of students in the class, indicators for the number of exams, and indicators for years in the program,  $\mathbf{z}$ . The factors are approximated by cross-sectional averages including the average grade for class  $c$  at year  $t$  and the average size of class  $c$  at year  $t$ .

The variable  $\mathbf{d}$  is a function of the vector of instruments  $\mathbf{w}$ . We take advantage of the experimental research design by using instruments generated from a random assignment of students into classes. For instance, while we have data on actual class size, we also have the number of students that were assigned to each class by the administration. We will consider class size, and the percentages of female students in a class and high income students in a class as endogenous variables. These variables will be instrumented with the number of students and the percentages of female students and high income students originally assigned by the lotteries that were used by the administration. The need for instrumentation arises because in spite of the administrative assignment of students to classes a small number of students will switch between classes, which may be indicative of a selection effect.

It is natural to control for individual and class-cohort latent heterogeneity by imposing a linear additive structure  $\lambda_i + f_{ct}$  as in Hanushek et al. (2003). The individual effect  $\lambda_i$ 's could be associated with motivation and ability to absorb knowledge when the student is listening to lectures or reading, and the factor  $f_{ct}$  could be interpreted as measuring teaching quality. In our specification however, the  $r$ -th component of the term  $\mathbf{f}'\boldsymbol{\lambda}$  could represent the interaction between students'  $i$  intrinsic motivation  $\lambda_{ir}$  and the quality of the teacher in a class  $f_{ctr}$ . High teaching quality may have a modest effect on the educational attainment of relatively unmotivated students, although it may dramatically affect performance among strong, motivated students. It is also natural to think that class size, teaching quality and student motivation are not stochastically independent. Notice that if student's motivation and teacher's quality enter multiplicatively in the scholastic



attainment function, the standard least squares transformations designed to remove time invariant heterogeneity produce biased results.

The remaining term  $u$  represents random shocks affecting a student's academic performance. At the same time, these shocks could also affect the number of students in a class, because they may affect a student's decision to stay in the assigned class. We model these potential interactions by letting the error term  $u$  in the educational achievement production function to be correlated with  $v$ .

#### 6.4. Empirical Analysis

We apply the method proposed in equations 3.3-3.7 to the structural model 6.1-6.2. Figure 6.1 presents results for the effects of interest. The figure presents estimates of the effects of the main covariates as a function of the quantiles  $\tau$  of the conditional distribution of educational attainment. The figure indicates that a reduction of the size of a class has a beneficial effect on academic achievement at the lowest quantiles, and an adverse effect on academic achievement at the highest quantiles. This suggests that a reduction of the size of a class can improve the achievement of low-performers, while reducing the achievement of high-performers. This evidence is silent on why we find heterogeneous class-size effects, but we offer a few possible channels. Smaller classes may allow weak students to interact more easily with instructors. If teachers are interested in increasing their mean evaluation, it is possible that they are more effectively targeting instructional resources including tutorial sessions to the weakest students in class. On the other hand, if high ability students learn from their peers, a reduction of the size of the class may hurt their performance.

It is interesting to see that the mean effect incorrectly suggest that a reduction of class size has a small, insignificant effect on performance. Moreover, we observe that changes in the percentage of female students in class significantly affect performance at the tails of the conditional distribution. We see evidence of heterogeneous effects across quantiles, ranging from a positive estimated effect at the lower tail to a negative estimated effect at the upper tail. On the other hand, the effects of changes in the proportion of high-income students do not seem to affect performance. The results also suggest that female students perform better than male students, and that students who were high-performers in the entry test perform better than students who were low-performers.

Although the students were randomly assigned into classes, teachers were not. A legitimate concern, for instance, would be that the best teachers are assigned by the administration to teach small classes. This issue is analyzed in De Giorgi, Pellizari and Woolston (2009), who find evidence that

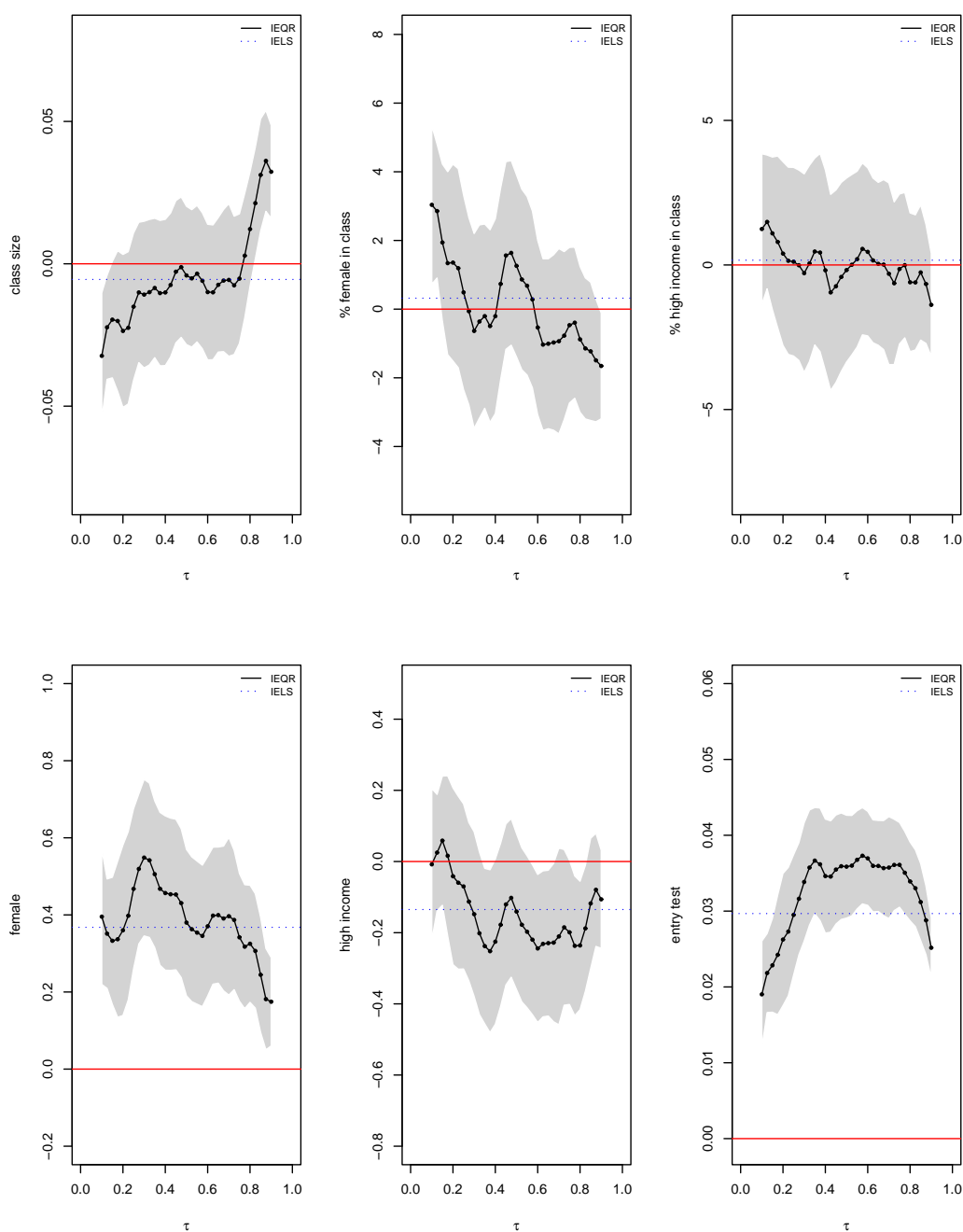


FIGURE 6.1. *Quantile regression covariate effects on educational attainment. The continuous dotted line shows quantile regression with interactive effects (QRIE) estimates and the dashed horizontal line shows the least squares version of the estimator (LSIE). These estimates were obtained from a sample of students in Economics, Management, and Finance.*

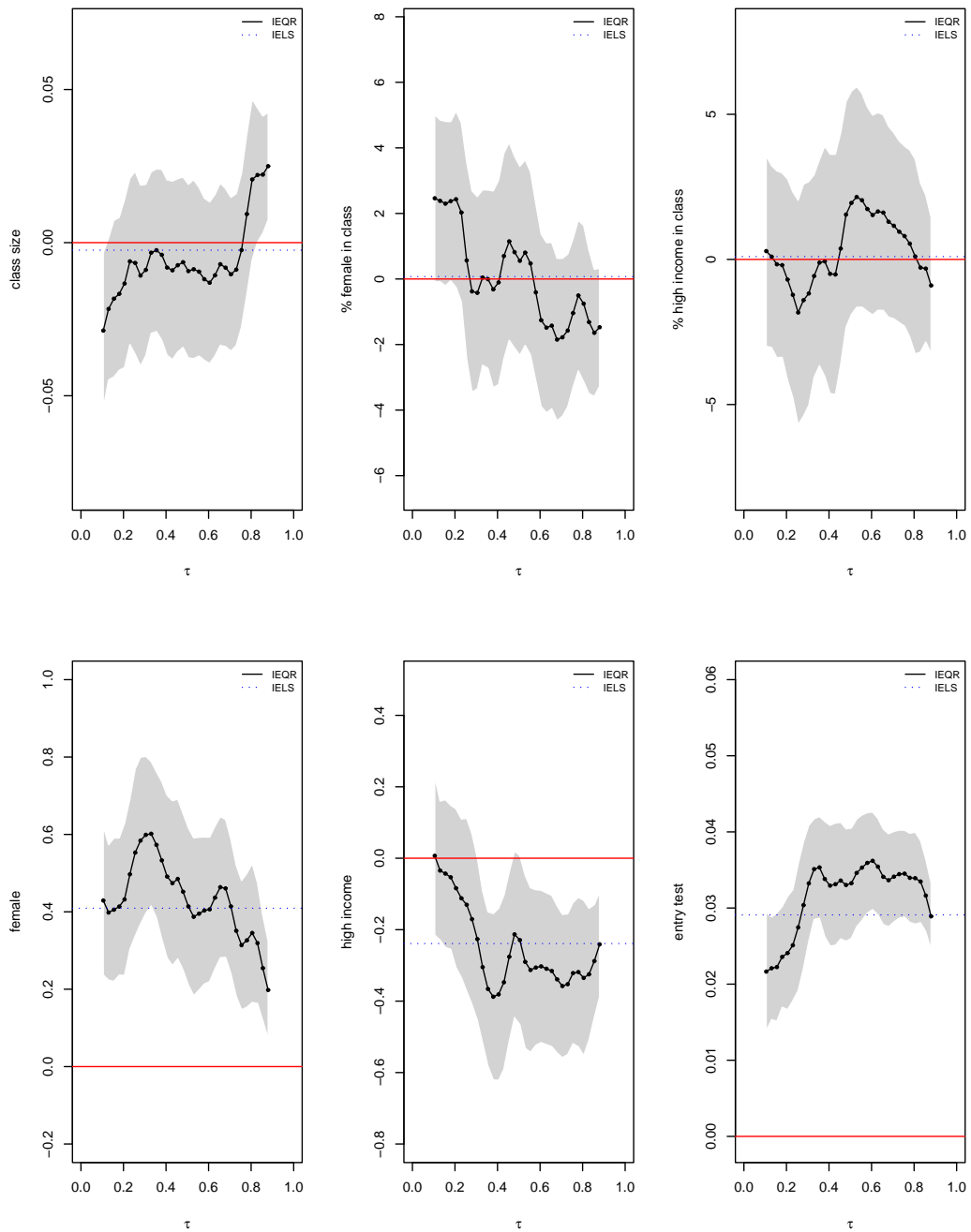


FIGURE 6.2. *Quantile regression covariate effects on educational attainment. The continuous dotted line shows quantile regression with interactive effects (QRIE) estimates and the dashed horizontal line shows the least squares version of the estimator (LSIE). These estimates were obtained from a sample of students in Management, and Finance.*

teachers' allocation is not a concern in this data. Using Figure 6.2, we investigate this potential issue by restricting the original sample to include classes of relatively similar size. By eliminating classes in the Economics program, we can reduce the standard deviation of class size from 16 students to 10 students and increase the minimum class-size from 90 students to 124 students (Table 6.1). We find that the effect of class size on academic performance suggests similar conclusions than before. We also observe that the effect of class composition measured by percentage of females in class and percentage of high-income students in class do not seem to significantly affect performance. The sole exception is the effect of percentage of females in class at the 0.9 quantile.

Let us now consider the empirical evidence presented in Table 6.2. This table presents the estimated effects of the variables that may be of interest to policy makers: class size and class composition, measured by the percentage of female students in class and the percentage of high income students in class. We present results from different methods including classical quantile regression (QR), instrumental variable quantile regression (QRIV), the fixed effects estimator (QRFE) developed by Koenker (2004), the instrumental variable estimator with fixed effects (QRIVFE) introduced by Harding and Lamarche (2009), and the method that uses instrumental variable estimation in a model with interactive effects (QRIE). The first five columns show the quantile regression version of the method presented in the first column, and the last column presents the corresponding least squares approach.

The pooled quantile method and the instrumental variable method suggest, in general, a negative effect of larger class sizes, although the only significant effects are at the 0.75 quantile. The results from a model with fixed effects indicate that all students benefit from a class size reduction with stronger students and weak students are similarly impacted by this policy. The findings associated with the class of instrumental variable-fixed effects estimator however indicate a different conclusion. The students that are benefiting from a reduction in the size of a class are the weak students, and the strong students are not affected by this policy. The results from the proposed approach suggest similar conclusions, but the effects are larger in absolute value at the tails. The evidence thus suggests that for weaker students, smaller classes are slightly better, and for the strongest students, smaller classes have a negative impact on academic achievement. This suggests that weaker students benefit from greater attention to their specific education needs and additional teacher attention in small classes. By contrast strong students may in fact benefit from large classes. It is possible that large classes facilitate numerous peer interactions and the increased competition between students acts as a strong motivational device. It is equally plausible though that in large classes strong

Method	Quantiles					Mean
	0.1	0.25	0.5	0.75	0.9	
Class size effect						
Pooled methods	-0.005 (0.004)	-0.001 (0.004)	0.003 (0.004)	-0.006 (0.003)	-0.004 (0.004)	-0.002 (0.003)
Instrumental Variable	-0.005 (0.004)	-0.001 (0.005)	-0.001 (0.005)	-0.007 (0.004)	-0.005 (0.004)	-0.003 (0.003)
Fixed Effects	-0.089 (0.007)	-0.088 (0.006)	-0.090 (0.005)	-0.088 (0.005)	-0.088 (0.006)	-0.093 (0.003)
Instrumental Variable	-0.014 (0.002)	-0.013 (0.003)	-0.001 (0.006)	0.006 (0.001)	0.005 (0.001)	-0.019 (0.007)
Fixed Effects	-0.034 (0.014)	-0.014 (0.016)	-0.005 (0.015)	-0.006 (0.014)	0.031 (0.010)	-0.005 (0.010)
% female in class						
Pooled methods	2.702 (1.628)	-0.426 (1.618)	0.732 (1.632)	-0.609 (1.173)	-1.302 (1.121)	-0.048 (1.008)
Instrumental Variable	0.873 (1.276)	-1.616 (1.225)	-1.177 (1.211)	-1.313 (0.922)	-2.007 (0.897)	-0.797 (1.049)
Fixed Effects	2.105 (1.984)	3.049 (1.571)	2.958 (1.268)	2.960 (1.148)	2.422 (1.527)	2.799 (0.504)
Instrumental Variable	2.762 (0.551)	2.926 (0.871)	3.341 (1.940)	2.435 (0.284)	2.352 (0.226)	2.850 (0.703)
Fixed Effects	2.886 (1.491)	0.163 (1.778)	1.260 (1.611)	-0.510 (1.379)	-1.624 (0.942)	0.320 (1.047)
% high income in class						
Pooled methods	-3.289 (1.464)	-6.530 (1.503)	-6.931 (1.600)	-4.968 (1.103)	-4.498 (1.275)	-5.234 (1.022)
Instrumental Variable	-1.725 (1.831)	-4.838 (1.635)	-6.388 (1.765)	-4.231 (1.346)	-4.003 (1.383)	-4.095 (1.079)
Fixed Effects	3.798 (1.947)	4.187 (1.582)	3.626 (1.342)	3.701 (1.384)	3.420 (1.765)	4.273 (0.657)
Instrumental Variable	4.897 (0.712)	5.111 (1.236)	5.039 (2.835)	4.289 (0.435)	4.181 (0.328)	5.504 (0.973)
Fixed Effects	1.111 (1.661)	0.115 (2.016)	-0.186 (1.926)	0.038 (1.581)	-1.402 (1.096)	0.165 (1.240)

TABLE 6.2. *Panel data and instrumental variable estimates of the causal effect of class size and class observed heterogeneity on educational achievement. Standard errors are presented in parentheses.*

students tend to monopolize and dominate their environment while weak students are less engaged in the class. Thus strong students end up benefiting more in such settings.

Effects of Interest	Quantiles					Pair of quantiles: 0.1 and			
	0.1	0.25	0.5	0.75	0.9	0.25	0.50	0.75	0.90
Specification I: Equations 6.1-6.2									
Class size effect	0.018	0.390	0.730	0.659	0.002	0.371	0.163	0.168	0.000
% Female in class	0.008	0.889	0.208	0.578	0.011	0.090	0.273	0.017	0.000
% High income in class	0.445	0.944	0.905	0.978	0.146	0.649	0.543	0.590	0.150
Interactive effects	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Specification II: Model with exogenous covariates and classical individual effects									
Class size effect	0.227	0.224	0.109	0.578	0.573	0.707	0.867	0.438	0.548
% Female in class	0.568	0.319	0.138	0.345	0.361	0.816	0.728	0.975	0.886
% High income in class	0.919	0.623	0.344	0.166	0.277	0.726	0.584	0.433	0.511
Interactive effects	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 6.3. *Specification tests in a quantile regression model with interactive effects. The interactive effects term includes cross-sectional averages of grade, class size and enrollment. The model with individual effects is estimated without time-invariant covariates. The table presents p-values.*

The results of Table 6.2 seem to suggest again that ignoring endogeneity and individual heterogeneity could lead to incorrect policies. The classical quantile regression estimator and the instrumental variable estimator suggest that an increase in the percentage of high-income students in a class may decrease the educational attainment of weak and strong students. The evidence associated with the fixed effects approaches suggests positive and significant effects, which contradicts the results obtained from our interactive approach. Changing the percentage of high income students in class does not seem to be associated with achievement gains.

## 6.5. Hypothesis Tests in a Model with Interactive Effects

In this Section, we briefly test if class size, the percent of females in class, and the percent of high income students in class are significantly different across quantiles. We test two hypotheses. First, we check whether the effects of interest are significantly different than zero at several quantiles of the conditional distribution of educational attainment. We also test if the interactive effect term is statistically significantly different than zero. Second, we test if the effects at the 0.1 quantile are significantly different than effects at other quantiles  $\{0.25, 0.50, 0.75, 0.90\}$ . These specification tests are performed on two models. First, we consider the model described in equations 6.1 and 6.2.

We also introduce a variation to equation 6.1, assuming that the covariates are exogenous and that the model includes both fixed and interactive effects. This allows us to examine the importance of the interactive effect term in the classical fixed effects model for educational attainment.

The results of several tests are presented in Table 6.3. While the upper block shows p-values of tests performed on our basic specification, the lower block presents p-values of tests performed on the variation of the model. We find that size effect are statistically different at the lower and upper tails, suggesting that changes in class size have a significant effect among low and high performers. We also see that changes in the gender composition of the class impact significantly different low- and high- performers. Moreover, it is interesting to see that the Wald tests reject the null hypothesis of no interactive effects in the model with fixed effects (p-values are 0.000). We interpret this evidence as suggesting that the fixed effects specification is rejected in favor of the interactive effects specification.

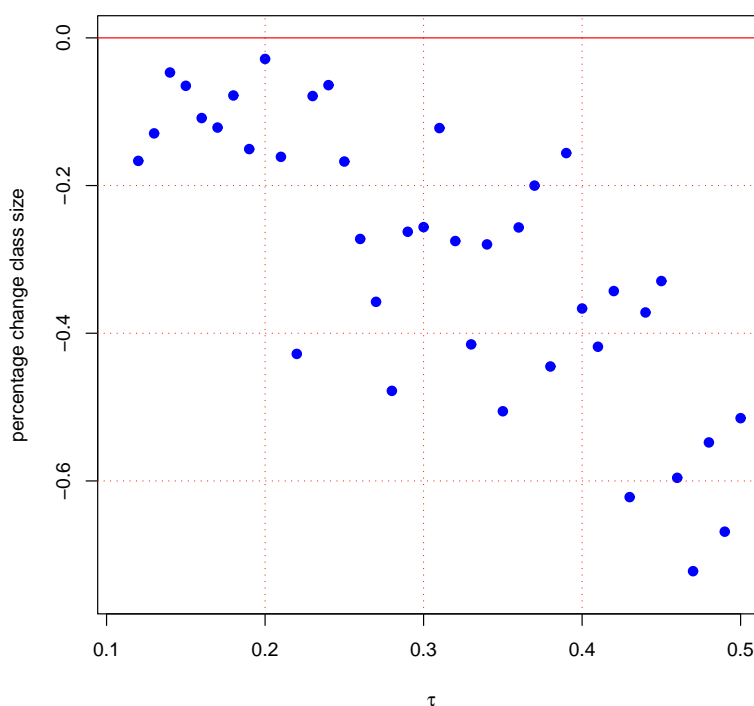


FIGURE 6.3. *Class size changes to prevent failure among the worst students.*

## 6.6. Improving Weak Students Performance

The previous analysis shows that for the poorly performing students, smaller classes appear to be improving performance. In this section, we briefly investigate what is the class size reduction needed to increase the achievement of these students at the bottom  $\tau_L$  percentile in the conditional distribution, to be at  $\tau'$ , with  $\tau' > \tau_L$ . In the case of two quantiles, the problem can be simply formulated as minimizing  $y(\tau_H) - y(\tau_L)$  subject to,

$$y(\tau_j) = q_{-c}(\tau_j) + \alpha_c(\tau_j)d_c,$$

for  $\tau_j = \{\tau_L, \tau_H\}$ . The function  $q_{-c}$  denotes the conditional quantile model corresponding to equation 6.1 that leaves out the term  $\alpha_c(\tau_j)d_c$ . The variable  $d_c$  is the number of students in class  $c$  and  $\alpha_c(\tau_j)$  is the class-size effect at the quantile  $\tau_j$ . An additional constraint is that the total number of students  $S$  is split for simplicity in two classes,  $c$  and  $c'$ . It is straightforward to show that the solution has the following form:

$$d_c^* = \frac{(q_{-c}(\tau_H) - q_{-c}(\tau_L)) + \alpha_c(\tau_H)S}{\alpha_c(\tau_H) + \alpha_c(\tau_L)}.$$

The implications of this expression are intuitive. Consider for simplicity the case of  $q_{-c}(\tau_H) = q_{-c}(\tau_L)$ . The solution  $d_c^*$  suggests to have a class with few students if a reduction in the size of the class does not significantly impact achievement at the highest quantile. If the effect of class size does not change across the quantiles of educational attainment, the solution  $d_c^*$  indicates to split the students in classes of equal size.

Using Figure 6.3, we construct the percentage change in class size needed to increase the achievement of the weak-performing students at the bottom 10 percentile in the conditional distribution, to be at  $\tau = \{0.11, 0.12, \dots, 0.5\}$ . For instance, the first point on the left represents the percentage change in class size  $d_c$  needed to increase the students' score from the bottom 10 to the bottom 11 percentile of the conditional educational attainment distribution. We consider  $S = 268$ , which is approximately twice the size of the average class in Table 6.1. We observe that most of the point estimates are contained in the square in the left corner, suggesting that a small percentage reduction of about 10 percent can positively impact the worst students in class. We also observe that the square in the top right corner of the plot is empty. This indicates that small reductions in the size of the class cannot produce academic gains comparable to the median student. Notice that making the weak performers comparable to the median performers does not seem to represent a feasible policy, because the size of the class has to be reduced by 60 percent.



The estimation of effects at the lower tail of the conditional distribution provides a convenient tool for the analysis of educational policies. Of course, one has to be cautious and interpret these results within the context of conditional quantiles. The analysis of the effect of educational policies on the unconditional attainment distribution is out of the scope of this paper and remains an open question within the panel data quantile regression literature.

## 7. Conclusion

This paper proposes a quantile regression estimator for a model with interactive effects potentially correlated with the independent variables and endogenous treatment effects. We provide conditions under which the slope parameter estimator is consistent and asymptotically Gaussian. Monte Carlo studies are carried out to investigate the finite sample performance of the proposed method in comparison with other candidate methods. The evidence shows that the finite sample performance of the proposed method is excellent under different Monte Carlo designs. We also apply the new method to an investigation of the effect of class size on educational performance.

Several directions remain to be investigated. Inferential procedures could be implemented by accommodating standard approaches, but they require a detailed investigation in the class of models with interactive effects. Moreover, the presence of a large number of loadings and factors suggests an attractive setting for regularization, which could represent an effective procedure to simultaneously improve the performance of the method and do model selection.

## Appendix A. Proofs

*Proof of Proposition 1.* The estimator of  $\boldsymbol{\vartheta}$  is equal to  $\hat{\boldsymbol{\vartheta}} = (\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{y})$ , or, equivalently,

$$\hat{\boldsymbol{\vartheta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}} \\ \hat{\boldsymbol{\delta}}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{D}'\tilde{\mathbf{P}}\mathbf{D} & \mathbf{D}'\tilde{\mathbf{P}}\tilde{\mathbf{D}} \\ \tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{D} & \tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{D}'\tilde{\mathbf{P}}\mathbf{y} \\ \tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{y} \end{pmatrix}$$

Standard partitioned matrix formulas give,

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{D}'\tilde{\mathbf{P}}\mathbf{D} - \mathbf{D}'\tilde{\mathbf{P}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{P}}\mathbf{y} \\ &\quad - (\mathbf{D}'\tilde{\mathbf{P}}\mathbf{D} - \mathbf{D}'\tilde{\mathbf{P}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{P}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{y} \\ &= (\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{P}}\mathbf{y} - (\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{P}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\mathbf{y} \end{aligned}$$

where  $\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}} = \tilde{\mathbf{P}} - \tilde{\mathbf{P}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}'\tilde{\mathbf{P}}\tilde{\mathbf{D}})^{-1}\tilde{\mathbf{D}}'\tilde{\mathbf{P}}$ . Therefore, we have that,

$$\hat{\boldsymbol{\alpha}} = (\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{y}.$$

Replacing  $\mathbf{y}$  by equation 2.3, we obtain,

$$(A.1) \quad \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{F}\boldsymbol{\lambda} + (\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{D})^{-1}\mathbf{D}'\tilde{\mathbf{M}}_{\tilde{\mathbf{P}}}\mathbf{u},$$

because  $\tilde{\mathbf{P}}\mathbf{X} = \mathbf{0}$ . After properly normalizing each matrix in A.1 by  $(NT)^{-1}$ , we have large  $N$ , large  $T$  convergence results. By Pesaran's (2006) Lemmas 2 and 3 and Assumptions 1-5, the first term is asymptotically negligible. The second term converges to zero by Assumption 4 and standard IV arguments.  $\square$

*Proof of Proposition 2.* The proofs follow closely the arguments found in Chernozhukov and Hansen (2008, Proposition 2), Galvao (2009, Theorem 2), and Lamarche (2010, Theorem 1).

(Consistency) Under the regularity conditions, identification and consistency results immediately follows from the proof of Theorem 3 in Chernozhukov and Hansen (2006) and Corollary 3.2.3 in van der Vaart and Wellner (1996). By proposition 2 in Chernozhukov and Hansen (2008), we have that  $\sup_{\boldsymbol{\alpha} \in \mathcal{A}} \|\hat{\boldsymbol{\vartheta}}(\boldsymbol{\alpha}, \tau) - \boldsymbol{\vartheta}(\boldsymbol{\alpha}, \tau)\| \rightarrow 0$  for  $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \boldsymbol{\gamma}')'$ . This implies that  $\sup_{\boldsymbol{\alpha} \in \mathcal{A}} \|\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}, \tau) - \boldsymbol{\gamma}(\boldsymbol{\alpha}, \tau)\| \rightarrow 0$ , and that  $\|\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}(\tau)\| \rightarrow 0$ . Consider a small ball  $\boldsymbol{\alpha}_n$  of radius  $r_n$  centered at  $\boldsymbol{\alpha}(\tau)$ . Then for any  $\boldsymbol{\alpha}_n \rightarrow \boldsymbol{\alpha}(\tau)$ , we have that  $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \tau) \rightarrow \boldsymbol{\beta}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\beta}(\tau)$ ,  $\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha}_n, \tau) \rightarrow \boldsymbol{\delta}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\delta}(\tau)$ , and  $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_n, \tau) \rightarrow \boldsymbol{\gamma}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\gamma}(\tau) = \mathbf{0}$ . Hence  $\hat{\boldsymbol{\vartheta}}(\boldsymbol{\alpha}_n, \tau) \rightarrow \boldsymbol{\vartheta}(\boldsymbol{\alpha}(\tau), \tau)$  for any  $\boldsymbol{\alpha}_n \rightarrow \boldsymbol{\alpha}(\tau)$ .

(Asymptotic Normality) For any  $\boldsymbol{\alpha}_n$ , we can write  $\rho_\tau(y_{it} - \mathbf{d}'_{it}\hat{\boldsymbol{\alpha}}(\tau) - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}(\tau) - \mathbf{f}'_t\hat{\boldsymbol{\delta}}(\tau) - \mathbf{w}'_{it}\hat{\boldsymbol{\gamma}}(\tau))$  as  $\rho_\tau(y_{it} - \xi_{it}(\tau) - \mathbf{d}'_{it}\hat{\boldsymbol{\delta}}_\alpha/\sqrt{NT} - \mathbf{x}'_{it}\hat{\boldsymbol{\delta}}_\beta/\sqrt{TN} - \mathbf{f}'_t\hat{\boldsymbol{\delta}}_\lambda/\sqrt{NT} - \mathbf{w}'_{it}\hat{\boldsymbol{\delta}}_\gamma/\sqrt{NT})$ , where  $\xi_{it}(\tau) = \mathbf{d}'_{it}\boldsymbol{\alpha}(\tau) + \mathbf{x}'_{it}\boldsymbol{\beta}(\tau) + \mathbf{f}'_t\boldsymbol{\delta}(\tau) + \mathbf{w}'_{it}\boldsymbol{\gamma}(\tau)$ ,  $\hat{\boldsymbol{\delta}}_\alpha(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\alpha}(\tau))$ ,  $\hat{\boldsymbol{\delta}}_\beta(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\beta}(\tau))$ ,  $\hat{\boldsymbol{\delta}}_\lambda(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\delta}(\tau))$ , and  $\hat{\boldsymbol{\delta}}_\gamma(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_n, \tau) - \mathbf{0})$ . Under assumption 4, the solution of 3.6 is equivalent to the solution of minimizing,

$$V_{TN}(\boldsymbol{\delta}) = \sum_{t=1}^T \sum_{i=1}^N \rho_\tau \left( u_{it}(\tau) - \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} - \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{NT}} - \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) - \rho_\tau(u_{it}(\tau))$$

where  $u_{it}(\tau) = y_{it} - \xi_{it}(\tau)$ . Following the conditions and argument of Ruppert and Carroll (1980), and Koenker and Portnoy (1987),

$$(A.2) \quad \sup \|v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}) - \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}))\| = o_p(1)$$

where  $\|\cdot\|$  denotes the standard Euclidean norm of a vector,  $\psi_\tau(u) = \tau - I(u < 0)$ , and,

$$v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) = \frac{-1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}'_t \psi_\tau \left( u_{it}(\tau) - \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} - \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{NT}} - \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right)$$

Taking expectation and expanding  $v$  under condition 6, we obtain

$$\begin{aligned}
& \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})) = \\
& = -\mathbb{E}\left(\frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau \left( y_{it} - \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} - \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{TN}} - \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right)\right) \\
& \quad + \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau)) \\
& = -\frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \left( G_{it} \left( \xi_{it}(\tau) + \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{TN}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) - \tau \right) \\
& = -\frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t g_{it}(\xi_{it}(\tau)) \left( \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{NT}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) + o(1)
\end{aligned}$$

where  $G(\cdot)$  is the conditional distribution of  $y$ . Clearly,  $v(\hat{\boldsymbol{\delta}}_\alpha, \hat{\boldsymbol{\delta}}_\beta, \hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\lambda) \rightarrow 0$ , and thus  $\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})) = v(\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0})$ . This last expression can be written as,

$$\begin{aligned}
& \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t g_{it}(\xi_{it}(\tau)) \left( \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{TN}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) = \\
& = \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau))
\end{aligned}$$

Letting  $\tilde{\mathbf{f}} = \sum_{i=1}^N \sum_{t=1}^T g_{it}(\xi_{it}(\tau)) \mathbf{f}_t \mathbf{f}'_t$  and solving for  $\boldsymbol{\delta}_\lambda$ , we have,

$$\begin{aligned}
\mathbf{f}'_t \frac{\boldsymbol{\delta}_\lambda}{\sqrt{TN}} & = \mathbf{f}'_t \tilde{\mathbf{f}}^{-1} \left( -\sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t g_{it}(\xi_{it}(\tau)) \left( \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) \right. \\
& \quad \left. + \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau)) \right) + \frac{R_{it}}{\sqrt{TN}} \\
& = -\tilde{\mathbf{d}}_t(\tau)' \frac{\boldsymbol{\delta}_\alpha}{\sqrt{TN}} - \tilde{\mathbf{x}}_t(\tau)' \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \tilde{\mathbf{w}}_t(\tau)' \frac{\boldsymbol{\delta}_\gamma}{\sqrt{TN}} + \tilde{\mathbf{f}}^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau)) + \frac{R_{it}}{\sqrt{NT}}
\end{aligned}$$

where for instance  $\tilde{\mathbf{d}}_t(\tau) = \mathbf{f}'_t \tilde{\mathbf{f}}^{-1} \sum_{i=1}^N \sum_{t=1}^T g_{it}(\xi_{it}(\tau)) \mathbf{f}_t \mathbf{d}_{it}$ , and  $R_{it}$  is the remainder term. Substituting the  $\hat{\boldsymbol{\lambda}}$ 's we denote,

$$v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) = \frac{-1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \mathbf{h}_{it} \psi_\tau \left( u_{it}(\tau) - \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} - \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \mathbf{f}'_t \frac{\hat{\boldsymbol{\delta}}_\lambda}{\sqrt{NT}} - \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right)$$

where  $\mathbf{h}_{it} = (\mathbf{x}'_{it}, \mathbf{w}'_{it})'$ . By uniformity,

$$(A.3) \quad \sup \|v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}) - \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0}, \mathbf{0}, \mathbf{0}))\| = o_p(1)$$

Expanding as above we obtain

$$\begin{aligned}
& \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0}, \mathbf{0}, \mathbf{0})) = \\
& = -\mathbb{E}\left(\frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} \psi_\tau \left( y_{it} - \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} - \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \mathbf{f}'_t \frac{\hat{\boldsymbol{\delta}}_\lambda}{\sqrt{T}} - \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) \right. \\
& \quad \left. + \frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} \psi_\tau(y_{it} - \xi_{it}(\tau)) \right) \\
& = -\frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} \left( G_{it} \left( \xi_{it}(\tau) + \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{f}'_t \frac{\hat{\boldsymbol{\delta}}_\lambda}{\sqrt{NT}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right) - \tau \right) \\
& = -\frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} g_{it}(\xi_{it}(\tau)) \left( \mathbf{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \mathbf{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \mathbf{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right. \\
& \quad \left. - \tilde{\mathbf{d}}'_t(\tau) \frac{\boldsymbol{\delta}_\alpha(\tau)}{\sqrt{TN}} - \tilde{\mathbf{x}}'_t(\tau) \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} - \tilde{\mathbf{w}}'_t(\tau) \frac{\boldsymbol{\delta}_\gamma}{\sqrt{TN}} + \tilde{\mathbf{f}}^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau)) + \frac{R_{it}}{\sqrt{NT}} \right)
\end{aligned}$$

Notice that  $v(\hat{\boldsymbol{\delta}}_\alpha, \hat{\boldsymbol{\delta}}_\beta, \hat{\boldsymbol{\delta}}_\gamma) \rightarrow 0$ , and thus  $\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0}, \mathbf{0}, \mathbf{0})) = v(\mathbf{0}, \mathbf{0}, \mathbf{0})$ . Letting  $\boldsymbol{\delta}_\vartheta = (\boldsymbol{\delta}'_\beta, \boldsymbol{\delta}'_\gamma)'$ , we write the last expression as,

$$\begin{aligned}
& \frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} g_{it} \left( (\mathbf{d}'_{it} - \tilde{\mathbf{d}}'_t(\tau)) \frac{\boldsymbol{\delta}_\alpha(\tau)}{\sqrt{NT}} + (\mathbf{h}'_{it} - \tilde{\mathbf{h}}'_t(\tau)) \frac{\boldsymbol{\delta}_\vartheta(\tau)}{\sqrt{TN}} + \tilde{\mathbf{f}}^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbf{f}_t \psi_\tau(y_{it} - \xi_{it}(\tau)) \right) = \\
& \qquad \qquad \qquad \frac{1}{\sqrt{TN}} \sum_{t=1}^T \sum_{i=1}^N \mathbf{h}_{it} \psi_\tau(y_{it} - \xi_{it}(\tau)) - \frac{R_{it}}{\sqrt{NT}}
\end{aligned}$$

Alternatively, using more convenient notation, we write the last expression as,

$$\mathbf{J}_\alpha \boldsymbol{\delta}_\alpha + \mathbf{J}_\vartheta \boldsymbol{\delta}_\vartheta = \mathbb{J}_\psi - \mathbf{R}$$

where  $\mathbf{J}_\alpha = \lim_{N,T \rightarrow \infty} \tilde{\mathbf{H}}' \mathbf{M}'_F \boldsymbol{\Phi} \mathbf{M}_F \mathbf{D}$ ,  $\mathbf{J}_\vartheta = \lim_{N,T \rightarrow \infty} \tilde{\mathbf{H}}' \mathbf{M}'_F \boldsymbol{\Phi} \mathbf{M}_F \tilde{\mathbf{H}}$ , and  $\mathbb{J}_\psi$  is a mean zero random variable with covariance  $\tau(1-\tau) \tilde{\mathbf{H}}' \mathbf{M}'_F \mathbf{M}_F \tilde{\mathbf{H}}$ . The remainder term  $\mathbf{R}$  is  $o_p(1)$  under the regularity conditions. Letting  $[\bar{\mathbf{J}}'_\beta, \bar{\mathbf{J}}'_\gamma]'$  be a conformable partition of  $\mathbf{J}_\vartheta^{-1}$  as in Galvao (2009) and Chernozhukov and Hansen (2006), then,

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_\gamma &= \bar{\mathbf{J}}'_\gamma (\mathbb{J}_\psi - \mathbf{J}_\alpha \boldsymbol{\delta}_\alpha) \\
\hat{\boldsymbol{\delta}}_\beta &= \bar{\mathbf{J}}'_\beta (\mathbb{J}_\psi - \mathbf{J}_\alpha \boldsymbol{\delta}_\alpha)
\end{aligned}$$

Letting  $\mathbf{H} = \bar{\mathbf{J}}'_\gamma \mathbf{A} \bar{\mathbf{J}}_\gamma$  as in Chernozhukov and Hansen (2006), we have that  $\hat{\boldsymbol{\delta}}_\alpha = (\mathbf{J}'_\alpha \mathbf{H} \mathbf{J}_\alpha)^{-1} \mathbf{J}'_\alpha \mathbf{H} \mathbb{J}_\psi$ . Replacing it in the previous expression,

$$\hat{\boldsymbol{\delta}}_\gamma = \bar{\mathbf{J}}'_\gamma (\mathbb{J}_\psi - \mathbf{J}_\alpha \boldsymbol{\delta}_\alpha) = \bar{\mathbf{J}}'_\gamma (\mathbf{I} - \mathbf{J}_\alpha (\mathbf{J}'_\alpha \mathbf{H} \mathbf{J}_\alpha)^{-1} (\mathbf{J}'_\alpha \mathbf{H})) \mathbb{J}_\psi = \bar{\mathbf{J}}'_\gamma (\mathbf{I} - \mathbf{L}) \mathbb{J}_\psi = \bar{\mathbf{J}}'_\gamma \mathbf{M} \mathbb{J}_\psi$$

where  $\mathbf{L} = \mathbf{J}_\alpha [\mathbf{J}'_\alpha \mathbf{H} \mathbf{J}_\alpha]^{-1} \mathbf{J}'_\alpha \mathbf{H}$  and  $\mathbf{M} = \mathbf{I} - \mathbf{L}$ . Due to invertibility of  $\mathbf{J}_\alpha \bar{\mathbf{J}}_\gamma$ ,  $\hat{\boldsymbol{\delta}}_\gamma = \mathbf{0} \times O_p(1) + o_p(1)$ . Similarly, substituting back  $\boldsymbol{\delta}_\alpha$ , we obtain that  $\hat{\boldsymbol{\delta}}_\beta = \bar{\mathbf{J}}'_\beta (\mathbf{I} - \mathbf{L}) \mathbf{J}_\psi$ . By the regularity conditions, we have that,

$$\hat{\boldsymbol{\delta}} = \begin{pmatrix} \hat{\boldsymbol{\delta}}_\alpha(\boldsymbol{\alpha}_n, \tau) \\ \hat{\boldsymbol{\delta}}_\beta(\boldsymbol{\alpha}_n, \tau) \end{pmatrix} = \begin{pmatrix} \sqrt{TN}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\alpha}(\tau)) \\ \sqrt{TN}(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\beta}(\tau)) \end{pmatrix} \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{J}' \mathbf{S} \mathbf{J}).$$

□

## References

- ABREVAYA, J., AND C. DAHL (2008): “The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Regression Estimation on Panel Data,” *Journal of Business and Economics Statistics*, 26(4), 379–397.
- ANDO, T., AND R. S. TSAY (2010): “Quantile Regression Models with Factor-Augmented Predictors and Information Criterion,” *Econometrics Journal*, forthcoming.
- ANGRIST, J., AND V. LAVY (1999): “Using Maimonides’ Rule to Estimate The Effects of Class Size on Scholastic Achievement,” *Quarterly Journal of Economics*, 114, 533–575.
- BAI, J. (2009): “Panel Data Models with Interactive Fixed Effects,” *Econometrica*, 77(4), 1229–1279.
- BALTAGI, B. (2005): *Econometric Analysis of Panel Data*. Wiley, New York, 3rd edn.
- BANDIERA, O., V. LARCINESE, AND I. RASUL (2010): “Heterogeneous Class Size Effects: New Evidence from a Panel of University Students,” forthcoming, *Economic Journal*.
- CANAY, I. (2010): “A Note on Quantile Regression for Panel Data Models,” mimeo, Northwestern University.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND W. NEWEY (2009): “Quantile and Average Effects in Nonseparable Panel Models,” cemmap Working Papers, CWP29/09.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73(1), 245–262.
- (2006): “Instrumental quantile regression inference for structural and treatment effect models,” *Journal of Econometrics*, 132(2), 491–525.
- (2008): “Instrumental variable quantile regression: A robust inference approach,” *Journal of Econometrics*, 142(1), 379–398.
- COLEMAN, J. (1966): “Equality of Educational Opportunity,” Report for US Department of Health, Education, and Welfare, Office of Education, Washington D.C.
- DE GIORGI, G., M. PELLIZZARI, AND W. WOOLSTON (2009): “Class-size and Class Heterogeneity,” mimeo, Stanford University.
- DINARDO, J., AND D. S. LEE (2010): “Program Evaluation and Research Designs,” NBER WP 16016.

- DOKSUM, K. (1974): “Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case,” *Annals of Statistics*, 2, 267–277.
- DUFLO, E., P. DUPLAS, AND M. KREMER (2008): “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” NBER Working Paper 14475.
- GALVAO, A. (2009): “Quantile Regression for Dynamic Panel Data with Fixed Effects,” mimeo, University of Illinois at Urbana-Champaign.
- GRAHAM, B. S., J. HAHN, AND J. L. POWELL (2009): “The Incidental Parameter Problem in a Non-differentiable Panel Data Model,” *Economics Letters*, 105, 181–182.
- HANUSHEK, E. A., J. F. KAIN, J. M. MARKMAN, AND S. G. RIVKIN (2003): “Does Peer Ability Affect Student Achievement?,” *Journal of Applied Econometrics*, 18, 527–544.
- HARDING, M., AND C. LAMARCHE (2009): “A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables,” *Economics Letters*, 104, 133–135.
- HECKMAN, J. J., AND E. VYTLACIL (2001): “Policy-Relevant Treatment Effects,” *American Economic Review*, 91(2), 107–111.
- HOXBY, C. (2000): “The Effects of Class Size on Student Achievement: New Evidence from Population Variation,” *Quarterly Journal of Economics*, 115, 1239–1285.
- HSIAO, C. (2003): *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edn.
- KATO, K., AND A. F. GALVAO (2010): “Smoothed Quantile Regression for Panel Data,” mimeo, Hiroshima University.
- KOENKER, R. (2004): “Quantile Regression for Longitudinal Data,” *Journal of Multivariate Analysis*, 91, 74–89.
- (2005): *Quantile Regression*. Cambridge University Press.
- (2010): “Quantreg,” R package version 4.48, [www.r-project.org](http://www.r-project.org).
- KOENKER, R., AND G. BASSETT (1982): “Tests of Linear Hypothesis and  $\ell_1$  Estimation,” *Econometrica*, 50, 1577–1584.
- KOENKER, R., AND S. PORTNOY (1987): “L-estimation for Linear Models,” *Journal of the American Statistical Association*, 82, 851–857.
- KOENKER, R., AND Z. XIAO (2002): “Inference on the Quantile Regression Process,” *Econometrica*, 70, 1583–1612.
- KRUEGER, A. B. (1999): “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics*, 114, 497–532.
- LAMARCHE, C. (2010): “Robust Penalized Quantile Regression Estimation for Panel Data,” *Journal of Econometrics*, 157, 396–408.
- LEHMANN, E. (1974): *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day Inc., San Francisco, California.
- LEVIN, J. (2001): “For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement,” *Empirical Economics*, 26, 221–246.
- MA, L., AND R. KOENKER (2006): “Quantile Regression Methods for Recursive Structural Equation Models,” *Journal of Econometrics*, 134, 471–506.

- PESARAN, M. H. (2006): “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure,” *Econometrica*, 74(4), 967–1012.
- PONOMAREVA, M. (2010): “Quantile Regression for Panel Data Models with Fixed Effects and Small T: Identification and Estimation,” mimeo, University of Western Ontario.
- POWELL, D. (2009): “Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors,” Rand Working Paper WR-710.
- ROSEN, A. (2009): “Set Identification via Quantile Restrictions in Short Panels,” CEMMAP, working paper CWP 26/09.
- RUPPERT, D., AND R. CARROLL (1980): “Trimmed Least Squares Estimation in the Linear Model,” *Journal of the American Statistical Association*, 75, 828–838.
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer, New York.
- WEI, Y., AND X. HE (2006): “Conditional Growth Charts,” *Annals of Statistics*, 34(5), 2069–2097.