Quantile Regression Estimation of a Model with Interactive Effects[*]

Matthew Harding[†] and Carlos Lamarche[‡]

June 24, 2010

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. The paper presents an empirical application of the method to study the effect of class size and class composition on educational attainment. The findings suggest 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.

JEL: C23, C33

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

[*]We are grateful to seminar participants at Stanford University and the University of Oklahoma for useful comments, and to 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.

[†]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`

[‡]Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019; Phone: (405) 325-5857; Email: `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 $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + u_{it}$, where $\boldsymbol{\lambda}_i$ is an $r \times 1$ vector of factor loadings and $\boldsymbol{f}_t$ corresponds to the r common time-varying factors, and where both $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$ 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 $\boldsymbol{f}$ to be correlated with the independent variables. We also allow for the possibility that the covariate $\boldsymbol{x}$ 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. 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 $\boldsymbol{\lambda}$, $\boldsymbol{f}$ and u correlated with the independent variable $\boldsymbol{x}$. 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 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 effects (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. Graham, Hahn, and Powell (2009) show that there is no incidental parameter problem in a static, non-differentiable panel data model with exogenous regressors.

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:

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

$$\boldsymbol{d}_{it} = \boldsymbol{\Pi}_1'\boldsymbol{w}_{it} + \boldsymbol{\Pi}_2'\boldsymbol{x}_{it} + \boldsymbol{\Pi}_{3i}'\boldsymbol{f}_t + \boldsymbol{\Pi}_4\boldsymbol{\lambda}_i'\boldsymbol{f}_t + \boldsymbol{\Pi}_5'\boldsymbol{\lambda}_i + \boldsymbol{v}_{it} \tag{2.2}$$

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

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{F}\boldsymbol{\lambda} + \boldsymbol{u} \tag{2.3}$$

where $\boldsymbol{y}$ is an $NT \times 1$ vector, $\boldsymbol{D}$ is an $NT \times k_1$ matrix, $\boldsymbol{X}$ is an $NT \times k_2$ matrix, and $\boldsymbol{F}$ is an $NT \times r$ matrix. We let $\boldsymbol{W}$ 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 $\boldsymbol{v}$ 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 $\boldsymbol{v}$.

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

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

ASSUMPTION 3. *The factor loadings $\boldsymbol{\lambda_i} = \boldsymbol{\lambda} + \pi_i$ are distributed independently of u_{jt} and v_{jt} for all i and j with mean λ and finite variances.*

There is a natural connection between econometric approaches to factor models with latent variables and 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,

$$\boldsymbol{y}(\boldsymbol{\alpha}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{F}\boldsymbol{\lambda} + \boldsymbol{W}\boldsymbol{\gamma} + \boldsymbol{u} \tag{2.4}$$

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

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\operatorname{argmin}} \left\{\hat{\gamma}(\boldsymbol{\alpha})'\boldsymbol{W}'\boldsymbol{M}\boldsymbol{W}\hat{\gamma}(\boldsymbol{\alpha})\right\} = (\boldsymbol{D}'\boldsymbol{P}_{MW}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{P}_{MW}\boldsymbol{y}, \tag{2.5}$$

where $\boldsymbol{P}_{MW} = \boldsymbol{M}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{M}\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{M}$. In practice, this projection matrix 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,

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

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

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

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

ASSUMPTION 4. *The matrix $\bar{\boldsymbol{C}}_3 + \boldsymbol{C}_4\bar{\boldsymbol{\lambda}}$ converges to a limiting matrix $\tilde{\boldsymbol{C}}_3$ 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 $\bar{\boldsymbol{\xi}} = (\bar{\boldsymbol{\xi}}_1, \ldots, \bar{\boldsymbol{\xi}}_T)'$ and $\bar{\boldsymbol{\xi}}_t = N^{-1}\sum_{i=1}^{N}(\boldsymbol{\alpha}'\boldsymbol{v}_{it} + u_{it}, \boldsymbol{v}_{it})'$. 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 $\boldsymbol{\Gamma}_0 = \tilde{\boldsymbol{C}}_3(\tilde{\boldsymbol{C}}_3'\tilde{\boldsymbol{C}}_3)^{-1}$, $\boldsymbol{\Gamma}_1 = \boldsymbol{C}_1'\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_2 = \boldsymbol{C}_2'\boldsymbol{\Gamma}_0$, and $\boldsymbol{\Gamma}_5 = \boldsymbol{C}_5'\boldsymbol{\Gamma}_0$, we have that,

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

which suggest that the unknown factors can be approximated by $(\bar{\boldsymbol{z}}_t',\ \bar{\boldsymbol{x}}_t',\ \bar{\boldsymbol{w}}_t', 1)'$. The following proposition shows that if valid instruments $\boldsymbol{w}_{it}$ are present, we can perform an instrumental variable regression augmented by the 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 instrumental variable (IV) estimator of* $\boldsymbol{\alpha}$ *is equal to* $\hat{\boldsymbol{\alpha}} = (\boldsymbol{D}'\bar{\boldsymbol{P}}_{\bar{M}W}\boldsymbol{D})^{-1}(\boldsymbol{D}'\bar{\boldsymbol{P}}_{\bar{M}W}\boldsymbol{y})$, *where* $\bar{\boldsymbol{P}}_{\bar{M}W} = \bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}$, $\bar{\boldsymbol{M}} = \boldsymbol{I}_{NT} - \bar{\boldsymbol{H}}(\bar{\boldsymbol{H}}'\bar{\boldsymbol{H}})^{-1}\bar{\boldsymbol{H}}'$, $\bar{\boldsymbol{H}} = [\boldsymbol{X}, \bar{\boldsymbol{Z}}, \bar{\boldsymbol{X}}, \bar{\boldsymbol{W}}, \iota_{NT}]$.

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 his analysis accommodating to issues associated with the lack of independence between $\boldsymbol{d}$ and $(\boldsymbol{\lambda}', u)'$. 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 quantiles functions:

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

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

where τ is a quantile in the interval $(0,1)$ 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.

This 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 2009), the response variable y is educational attainment and is influenced by class size and peers effects $\boldsymbol{d}$, and individual, family and school characteristics $\boldsymbol{x}$. The last term u may represent idiosyncratic shocks to achievement that forces 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 students unobserved ability to absorb knowledge when listening to lectures, effort and motivation, and the variable $\boldsymbol{f}$ 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.

Remark 1. The conditional quantile function 3.1 allows for unobserved time-varying effects $\boldsymbol{f}_t$, 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 $\boldsymbol{v}$ in model 2.1-2.2 are stochastically independent. By setting $\boldsymbol{f}_t = r = 1$ for all t, we have the conditional quantile function $Q_{Y_{it}}(\tau|\boldsymbol{d}_{it}, \boldsymbol{x}_{it}, \lambda_i) = \boldsymbol{d}_{it}'\boldsymbol{\alpha}(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \lambda_i$ 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 $Q_{Y_{it}}(\tau|\boldsymbol{d}_{it}, \boldsymbol{x}_{it}, f_t) = \boldsymbol{d}_{it}'\boldsymbol{\alpha}(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + f_t$. A simple variation arises when $\boldsymbol{f}_t'\boldsymbol{\lambda}_i = \lambda_i + f_t$, which can be also 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).

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

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

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 $Q_{Y_{it}}(\tau|\boldsymbol{d}_{it}, \boldsymbol{x}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t)$ is evaluated at $\boldsymbol{d}_{it} = Q_{D_{it}}(\tau|\boldsymbol{w}_{it}, \boldsymbol{x}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t)$, we can substitute (3.2) into (3.1) and summing over the cross-sectional dimension of the model, we obtain,

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

where $\bar{\boldsymbol{z}}_t(\tau)$ is the cross-sectional average of $\boldsymbol{z}_{it}(\tau) = (Q_{Y_{it}}(\tau|\bullet), Q_{D_{it}}(\tau|\bullet)')'$, and $\boldsymbol{C}_2(\tau) = ((\boldsymbol{\alpha}'\boldsymbol{\Pi}_2' + \boldsymbol{\beta}(\tau)')', \boldsymbol{\Pi}_2')'$. As before we can augment the design matrix by adding the cross-sectional averages of the endogenous and exogenous variables, $\boldsymbol{f}_t(\tau) = \boldsymbol{\Psi}(\tau; \bar{\boldsymbol{z}}_t, \bar{\boldsymbol{w}}_t, \bar{\boldsymbol{x}}_t, 1)$. The second term $\boldsymbol{\Phi}_{it}(\tau) = \boldsymbol{\Phi}(\tau; \boldsymbol{w}_{it}, \boldsymbol{x}_{it}, \boldsymbol{f}_t, \boldsymbol{\lambda}_i)$ is a vector of transformations of instruments. It is possible to estimate $\boldsymbol{\Phi}$ by a least squares projection of the endogenous variables $\boldsymbol{d}$ on the instruments $\boldsymbol{w}$, the exogenous variables $\boldsymbol{x}$, and a vector of individual and time effects (see, e.g., Harding and Lamarche 2009). We consider the case of $\dim(\boldsymbol{\alpha}) = \dim(\boldsymbol{\gamma})$, although the vector $\boldsymbol{\Phi}$ may include more elements than the vector $\boldsymbol{d}$.

First we minimize the objective function above for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ as functions of τ and $\boldsymbol{\alpha}$,

$$\{\hat{\boldsymbol{\beta}}(\tau,\boldsymbol{\alpha}),\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha}),\hat{\boldsymbol{\delta}}(\tau,\boldsymbol{\alpha})\} = \underset{\beta,\gamma,\delta\in\mathcal{B}\times\mathcal{G}\times\mathcal{F}}{\operatorname{argmin}} \sum_{t=1}^{T}\sum_{i=1}^{N} C_{it}(\tau,\boldsymbol{\alpha};\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta}). \tag{3.5}$$

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

$$\hat{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\operatorname{argmin}} \left\{\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})'\hat{\boldsymbol{A}}(\tau)\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})\right\} \tag{3.6}$$

for a positive definite matrix $\boldsymbol{A}$. The parameter estimates are then given by,

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

Remark 2. 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 on the estimation procedure. Our approach is computationally simple, and we believe it is convenient for estimating large microeconometric 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) and Harding and Lamarche (2009). We define a grid $j = \{1,\ldots,J\}$. For a given j, we define a grid running ordinary quantile regression of $y_{it} - \boldsymbol{d}_{it}'\boldsymbol{\alpha}_j$ on $\boldsymbol{x}_{it}$, $\hat{\boldsymbol{f}}_t(\tau)$, 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 $\|\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})\|^2_{\hat{\boldsymbol{A}}(\tau)}$.

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 5. *The variables y_{it} are independent with conditional distribution G_{it}, and continuous densities g_{it} uniformly bounded away from 0 and ∞, with bounded derivatives g_{it}' everywhere.*

ASSUMPTION 6. *The variables $\boldsymbol{w}_{it}$ and u_{it} are stochastically independent and the number of endogenous variables k_1 is equal to the number of instruments m.*

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

$$\begin{aligned}\boldsymbol{\Pi}(\boldsymbol{\pi}, \tau) &\equiv \mathbb{E}\left[(\tau - 1\{Y < \boldsymbol{D\alpha} + \boldsymbol{X\beta} + \boldsymbol{W\gamma} + \boldsymbol{F\delta}\}\boldsymbol{\Delta}(\tau)\right] \\ \boldsymbol{\Pi}(\boldsymbol{\theta}, \tau) &\equiv \mathbb{E}\left[(\tau - 1\{Y < \boldsymbol{D\alpha} + \boldsymbol{X\beta} + \boldsymbol{F\lambda}\}\boldsymbol{\Delta}(\tau)\right]\end{aligned}$$

where $\boldsymbol{\Delta}(\tau) = [\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{F}]'$. *The Jacobian matrices* $\partial\boldsymbol{\Pi}(\boldsymbol{\theta}, \tau)/\partial(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\lambda}')$ *and* $\partial\boldsymbol{\Pi}(\boldsymbol{\pi}, \tau)/\partial(\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}')$ *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* $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}) \mapsto \boldsymbol{\Pi}(\boldsymbol{\theta}, \tau)$ *is simply connected.*

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

$$\begin{aligned}\boldsymbol{S}(\tau) &= \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{\tau(1-\tau)}{NT}\tilde{\boldsymbol{X}}'\boldsymbol{M}_F'\boldsymbol{M}_F\tilde{\boldsymbol{X}} \\ \boldsymbol{J}(\tau) &= \lim_{\substack{T\to\infty \\ N\to\infty}} \frac{1}{NT}(\boldsymbol{K}', \boldsymbol{L}')\end{aligned}$$

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

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

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 5 ensures a well-defined asymptotic behavior of the quantile regression estimator. Condition 6 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. Condition 10 is important both for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function.

THEOREM 1. *Under the regularity conditions, the estimator* $(\hat{\boldsymbol{\alpha}}(\tau)', \hat{\boldsymbol{\beta}}(\tau)')'$ *is consistent and asymptotically normally distributed with mean* $(\boldsymbol{\alpha}(\tau)', \boldsymbol{\beta}(\tau)')'$ *and covariance matrix* $\boldsymbol{J}'(\tau)\boldsymbol{S}(\tau)\boldsymbol{J}(\tau)$.

Estimation and basic inference requires us to approximate $\boldsymbol{F}$ and estimate the conditional density g. The matrix of factors $\boldsymbol{F}$ of dimension $NT \times r$ is replaced by a matrix $\hat{\boldsymbol{F}}$ 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 $\hat{\boldsymbol{F}}(\tau)$ 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 $\hat{u}_{it}(\tau) = y_{it} - \boldsymbol{d}'_{it}\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau) - \hat{\boldsymbol{f}}'_t\hat{\boldsymbol{\delta}}(\tau)$ and a properly chosen bandwidth h (see Koenker 2005 for details). Therefore, the matrix $\boldsymbol{S}(\tau)$ can be estimated by,

$$\hat{\boldsymbol{S}}(\tau) = \frac{\tau(1-\tau)}{NT}\tilde{\boldsymbol{X}}'\hat{\boldsymbol{M}}'_{\hat{F}}\hat{\boldsymbol{M}}_{\hat{F}}\tilde{\boldsymbol{X}}, \tag{4.1}$$

where $\tilde{\boldsymbol{X}} = [\boldsymbol{W}, \boldsymbol{X}]$ and $\hat{\boldsymbol{M}}_{\hat{F}} = \boldsymbol{I} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{\Phi}}\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\hat{\boldsymbol{\Phi}}$. The matrix $\boldsymbol{J}(\tau)$ can be similarly estimated.

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):

$$y_{it} = \beta_0 + \beta_1 d_{it} + \gamma x_t + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + (1 + h d_{it}) u_{it} \tag{5.1}$$

$$d_{it} = \pi_0 + \pi_1 w_{it} + \pi_2 x_t + \pi_3 f_{1t} + \pi_3 f_{2t} + \pi_4 \lambda_{1i} f_{1t} + \pi_4 \lambda_{2i} f_{2t} + \epsilon_i + v_{it} \tag{5.2}$$

$$f_{jt} = \rho_f f_{jt-1} + \eta_{jt} \tag{5.3}$$

$$\eta_{jt} = \rho_\eta \eta_{jt-1} + e_{jt} \tag{5.4}$$

for $j = \{1, 2\}, \ldots t = -49, \ldots 0, \ldots T$ in the last two equations. The random variables are $x_t \sim \mathcal{N}(0, 1)$, $\lambda_{i1}, \lambda_{i2} \sim \mathcal{N}(1, 0.2)$, and e, ϵ and w are Gaussian independent random variables. The error terms are $(u_{it}, v_{it})' \sim (0, \boldsymbol{\Omega})$, distributed either as Gaussian or t-student distribution with two degrees of freedom. The parameters are assumed to be: $\beta_0 = \pi_3 = 2$, $\beta_1 = \gamma = \pi_0 = \pi_1 = \pi_2 = 1$, $\rho_f = 0.90$, $\rho_\eta = 0.25$, and $\Omega_{11} = \Omega_{22} = 1$. We consider three designs for the location shift model $h = 0$:

Statistic	Sample Size		Panel Data Methods: Least Squares OLS	LSFE	LSIVFE	LSIE	Quantile Regression QR	QRFE	QRIVFE	QRIE
	N	T	Gaussian distributions							
Bias	50	4	0.2153	0.2777	-0.0144	0.0010	0.2173	0.2786	0.0085	-0.0009
RMSE	50	4	0.2525	0.3107	0.1527	0.0914	0.2568	0.3131	0.2537	0.1119
Bias	100	4	0.2144	0.2811	-0.1709	0.0010	0.2163	0.2804	0.1570	0.0002
RMSE	100	4	0.2456	0.3073	3.2011	0.0588	0.2481	0.3065	3.5380	0.0767
Bias	50	8	0.3305	0.3788	-0.0067	0.0022	0.3308	0.3782	-0.0001	0.0019
RMSE	50	8	0.3410	0.3855	0.1076	0.0639	0.3419	0.3850	0.2084	0.0775
Bias	100	8	0.3279	0.3774	-0.0080	-0.0003	0.3269	0.3744	-0.0058	-0.0013
RMSE	100	8	0.3378	0.3837	0.0732	0.0396	0.3373	0.3812	0.1539	0.0494
	N	T	t_2 distributions							
Bias	50	4	0.2218	0.2914	-0.0458	-0.0069	0.2288	0.2915	-0.0233	-0.0006
RMSE	50	4	0.2721	0.3418	0.4456	0.2834	0.2594	0.3215	0.4298	0.1298
Bias	100	4	0.2065	0.2699	-0.0161	-0.0093	0.2117	0.2739	-0.0273	-0.0017
RMSE	100	4	0.2472	0.3080	0.2505	0.2063	0.2449	0.3025	0.3406	0.0872
Bias	50	8	0.3291	0.3774	-0.0063	0.0088	0.3286	0.3730	-0.0156	-0.0025
RMSE	50	8	0.3450	0.3899	0.1816	0.1437	0.3421	0.3826	0.3363	0.0956
Bias	100	8	0.3304	0.3782	-0.0171	-0.0068	0.3322	0.3773	-0.0271	-0.0028
RMSE	100	8	0.3434	0.3877	0.1440	0.1145	0.3429	0.3844	0.2138	0.0656

TABLE 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.*

Design 1: The endogenous variable d is not correlated with the λ'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 λ's, and the error terms in equations 5.1 and 5.2 are not correlated. We assume $\pi_4 = 2$ and $\Omega_{12} = \Omega_{21} = 0$.

Statistic	Sample Size		Panel Data Methods: Least Squares OLS	LSFE	LSIVFE	LSIE	Quantile Regression QR	QRFE	QRIVFE	QRIE
	N	T	Gaussian distributions							
Bias	50	4	0.2202	0.2001	-0.0587	-0.0054	0.2187	0.2010	0.1899	-0.0077
RMSE	50	4	0.2256	0.2085	3.2855	0.0963	0.2251	0.2099	4.5368	0.1168
Bias	100	4	0.2180	0.2043	-0.0112	-0.0011	0.2180	0.2057	0.0198	-0.0056
RMSE	100	4	0.2234	0.2100	0.3486	0.0600	0.2237	0.2114	0.5483	0.0820
Bias	50	8	0.2350	0.2331	-0.0298	-0.0004	0.2354	0.2333	0.0505	-0.0015
RMSE	50	8	0.2360	0.2338	0.2635	0.0637	0.2365	0.2341	0.5935	0.0752
Bias	100	8	0.2359	0.2331	-0.0243	-0.0013	0.2352	0.2323	0.0014	-0.0023
RMSE	100	8	0.2367	0.2336	0.1554	0.0400	0.2361	0.2329	0.3880	0.0507
	N	T	t_2 distributions							
Bias	50	4	0.2193	0.2097	-0.0178	-0.0137	0.2218	0.2114	0.0039	-0.0148
RMSE	50	4	0.2359	0.2307	0.8507	0.2893	0.2281	0.2194	1.4281	0.1405
Bias	100	4	0.2144	0.1956	-0.1152	-0.0110	0.2166	0.2032	0.0346	-0.0048
RMSE	100	4	0.2236	0.2096	2.1972	0.2061	0.2230	0.2102	1.0249	0.0945
Bias	50	8	0.2356	0.2324	-0.0449	0.0065	0.2347	0.2317	-0.3670	-0.0071
RMSE	50	8	0.2388	0.2360	0.7388	0.1431	0.2364	0.2331	7.5819	0.0997
Bias	100	8	0.2362	0.2330	-0.0327	-0.0083	0.2361	0.2336	-0.0729	-0.0029
RMSE	100	8	0.2381	0.2349	0.2912	0.1152	0.2371	0.2343	0.7727	0.0670

TABLE 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.*

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 λ's as in the experiment carried out in Design 2.

Statistic	Sample Size		Panel Data Methods: Least Squares OLS	LSFE	LSIVFE	LSIE	Quantile Regression QR	QRFE	QRIVFE	QRIE
	N	T	Gaussian distributions							
Bias	50	4	0.3064	0.2562	-0.1157	-0.0140	0.3065	0.2598	0.2082	-0.0146
RMSE	50	4	0.3101	0.2577	4.4951	0.1051	0.3104	0.2620	5.3424	0.1410
Bias	100	4	0.3050	0.2564	-0.0134	-0.0060	0.3050	0.2584	0.0288	-0.0076
RMSE	100	4	0.3077	0.2571	0.2846	0.0643	0.3078	0.2594	0.5769	0.0817
Bias	50	8	0.2742	0.2533	-0.0379	-0.0078	0.2747	0.2542	0.0271	-0.0121
RMSE	50	8	0.2750	0.2536	0.2896	0.0714	0.2757	0.2545	0.6125	0.0899
Bias	100	8	0.2749	0.2531	-0.0281	-0.0044	0.2747	0.2535	0.0019	-0.0084
RMSE	100	8	0.2758	0.2532	0.1731	0.0460	0.2757	0.2537	0.4358	0.0571
	N	T	t_2 distributions							
Bias	50	4	0.3008	0.2546	0.0284	-0.0150	0.2997	0.2553	-0.0326	-0.0247
RMSE	50	4	0.3087	0.2649	1.2329	0.2282	0.3032	0.2588	1.0831	0.1510
Bias	100	4	0.3069	0.2565	-0.0396	-0.0050	0.3045	0.2575	0.0858	-0.0030
RMSE	100	4	0.3127	0.2630	0.6784	0.1465	0.3077	0.2594	2.1683	0.0999
Bias	50	8	0.2775	0.2544	-0.1189	0.0010	0.2760	0.2546	-0.3926	-0.0104
RMSE	50	8	0.2801	0.2570	1.9215	0.2044	0.2772	0.2552	7.3960	0.0962
Bias	100	8	0.2737	0.2521	-0.0262	0.0068	0.2737	0.2528	-0.0799	-0.0032
RMSE	100	8	0.2752	0.2532	0.2697	0.1176	0.2745	0.2531	0.8351	0.0666

Table 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.*

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

Tables 1-3 present the bias and root mean square error (RMSE) of the simulation experiments. While the upper panel of the tables presents results for the case that u and v are distributed as

Statistic	Sample Size		0.25 Quantile				0.75 Quantile			
			QR	QRFE	QRIVFE	QRIE	QR	QRFE	QRIVFE	QRIE
	N	T	Monte Carlo Design 1 - Gaussian distributions							
Bias	50	4	0.2677	0.2975	0.0089	0.0155	0.2241	0.2452	-0.0613	-0.0208
RMSE	50	4	0.3059	0.3337	0.4267	0.1388	0.2666	0.2865	0.4641	0.1408
Bias	100	4	0.2755	0.3061	0.0724	0.0165	0.2331	0.2546	-0.0574	-0.0155
RMSE	100	4	0.3073	0.3349	1.2909	0.0976	0.2661	0.2865	0.2669	0.0960
Bias	50	8	0.3833	0.3942	0.0279	0.0190	0.3407	0.3510	-0.0626	-0.0140
RMSE	50	8	0.3939	0.4033	0.3352	0.0977	0.3512	0.3598	0.3216	0.0980
Bias	100	8	0.3869	0.3962	0.0384	0.0201	0.3410	0.3520	-0.0696	-0.0148
RMSE	100	8	0.3972	0.4047	0.1782	0.0690	0.3509	0.3603	0.3012	0.0713
	N	T	Monte Carlo Design 2 - Gaussian distributions							
Bias	50	4	0.2655	0.2386	0.0377	0.0206	0.2033	0.1641	-0.1005	-0.0434
RMSE	50	4	0.2787	0.2554	1.0407	0.1717	0.2161	0.1864	1.2236	0.2001
Bias	100	4	0.2693	0.2444	0.0129	0.0259	0.2048	0.1673	-0.0687	-0.0327
RMSE	100	4	0.2800	0.2552	1.0646	0.1245	0.2146	0.1816	0.5873	0.1237
Bias	50	8	0.2727	0.2648	0.0675	0.0274	0.2152	0.1994	-0.1008	-0.0306
RMSE	50	8	0.2767	0.2684	0.9335	0.1203	0.2195	0.2038	1.0904	0.1288
Bias	100	8	0.2756	0.2670	0.0820	0.0305	0.2114	0.1976	-0.0529	-0.0330
RMSE	100	8	0.2796	0.2702	0.6258	0.0891	0.2153	0.2018	1.5374	0.0952
	N	T	Monte Carlo Design 3 - Gaussian distributions							
Bias	50	4	0.3212	0.3020	0.0337	0.0128	0.2524	0.2211	-0.1074	-0.0512
RMSE	50	4	0.3309	0.3158	1.0062	0.1814	0.2770	0.2573	1.1633	0.1967
Bias	100	4	0.3218	0.2980	0.0063	0.0225	0.2471	0.2132	-0.0727	-0.0368
RMSE	100	4	0.3298	0.3083	1.0843	0.1207	0.2686	0.2439	0.5658	0.1289
Bias	50	8	0.2986	0.2888	0.0463	0.0178	0.2354	0.2207	-0.1097	-0.0414
RMSE	50	8	0.3015	0.2911	1.0504	0.1220	0.2437	0.2284	1.0598	0.1372
Bias	100	8	0.2998	0.2908	0.0724	0.0265	0.2270	0.2127	-0.0534	-0.0347
RMSE	100	8	0.3023	0.2927	0.6441	0.0869	0.2355	0.2216	1.5337	0.0924

TABLE 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.*

Gaussian random variables ($\mathcal{N}(0,1)$), the lower panel of the tables presents results for the case of u and v distributed as t-student distribution with two degrees of freedom (t_2). We present results based on $N = \{50, 100\}$ 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 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 fixed effects 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).

Tables 1 and 2 provide evidence on 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 even larger 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 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 errors follow a t_2 distribution, while the quantile estimator QRIE continues to perform well with only a minor increase in MSE.

Table 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 4 presents the bias and root mean square error in the Gaussian case when β_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 0.6 percent to 5.1 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.

Overall, the finite sample performances of the methods for models with interactive effects very good in all the variants of the models considered in Tables 1-4. 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 be 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. 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 the drivers of 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 Lavy 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 have been conducted 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 (2009). 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 impact achievement in language and smaller classes impact achievement in mathematics. Finally, Bandiera, Larcinese, and Rasul (2009) employ standard quantile regression 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.

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 5. *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.*

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 educational attainment, background demographic and socioeconomic characteristics such us 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 5 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 (Di 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. 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 includes 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:

$$y_{ict} = \boldsymbol{d}'_{ct}\boldsymbol{\alpha} + \boldsymbol{x}'_i\boldsymbol{\beta} + \boldsymbol{f}'_{ct}\boldsymbol{\lambda}_i + u_{ict} \tag{6.1}$$

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

where y is the average grade of student i in a class c at year t, and $\boldsymbol{d}$ 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 $\boldsymbol{x}$ 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 us 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. 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 $\boldsymbol{d}$ is a function of the vector of instruments $\boldsymbol{w}$. 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 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.

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 λ_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 $\boldsymbol{f}'\boldsymbol{\lambda}$ could represent the interaction between student's i intrinsic motivation λ_{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 student's academic performance. At the same time, these shocks could also affect the number of students in a class, because they may affect 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. **An Empirical Analysis**

We apply the method proposed in equations 3.3-3.6 to the structural model 6.1-6.2. Figure 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 τ of the conditional distribution of educational attainment.

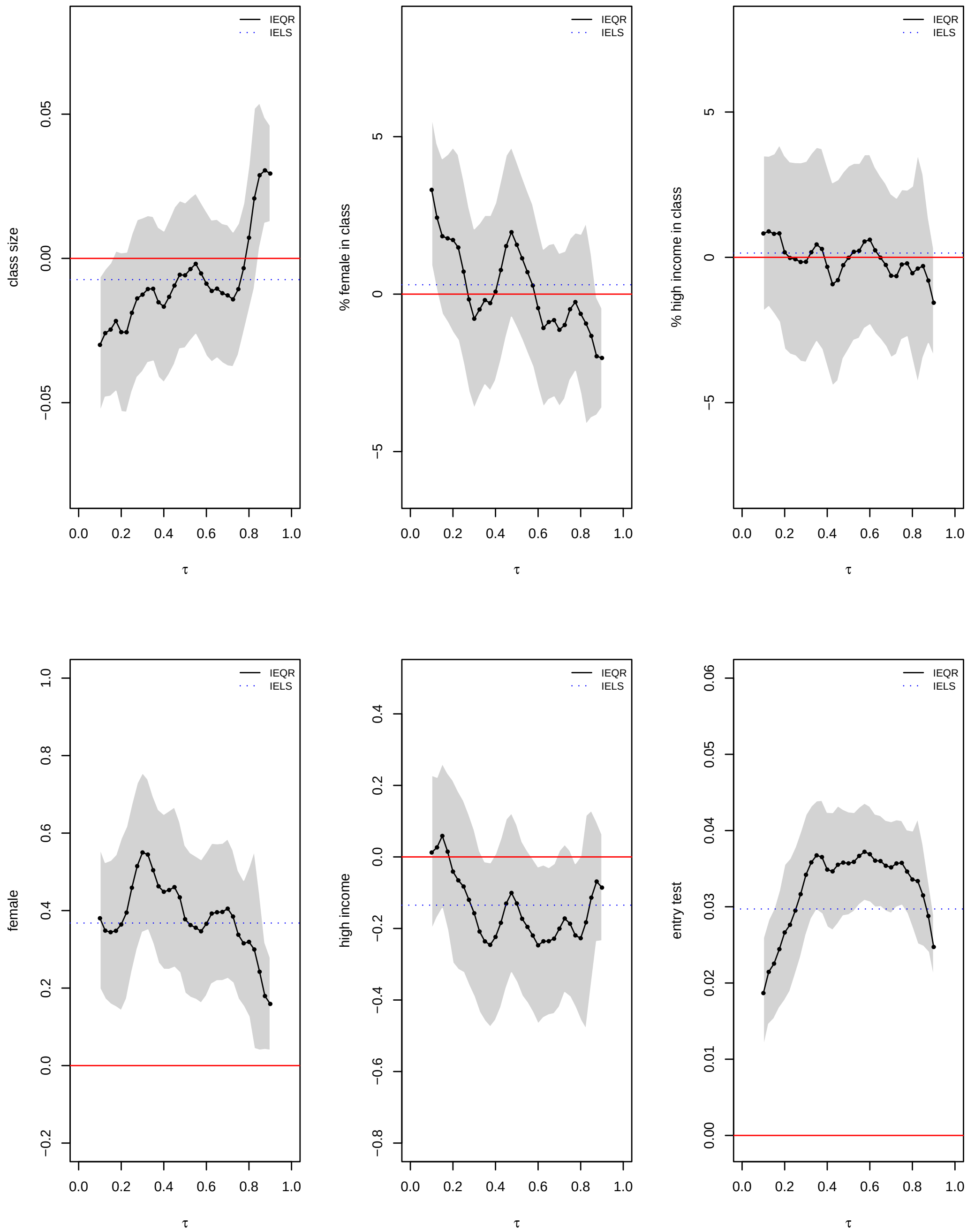

FIGURE 1. *Quantile regression covariate effects on educational attainment. The continuous dotted line shows quantile regression with interactive effects (QRIE) estimates and the dashed horizontal line shows the least squares version of the estimator (LSIE). These estimates were obtained from a sample of students in Economics, Management, and Finance.*

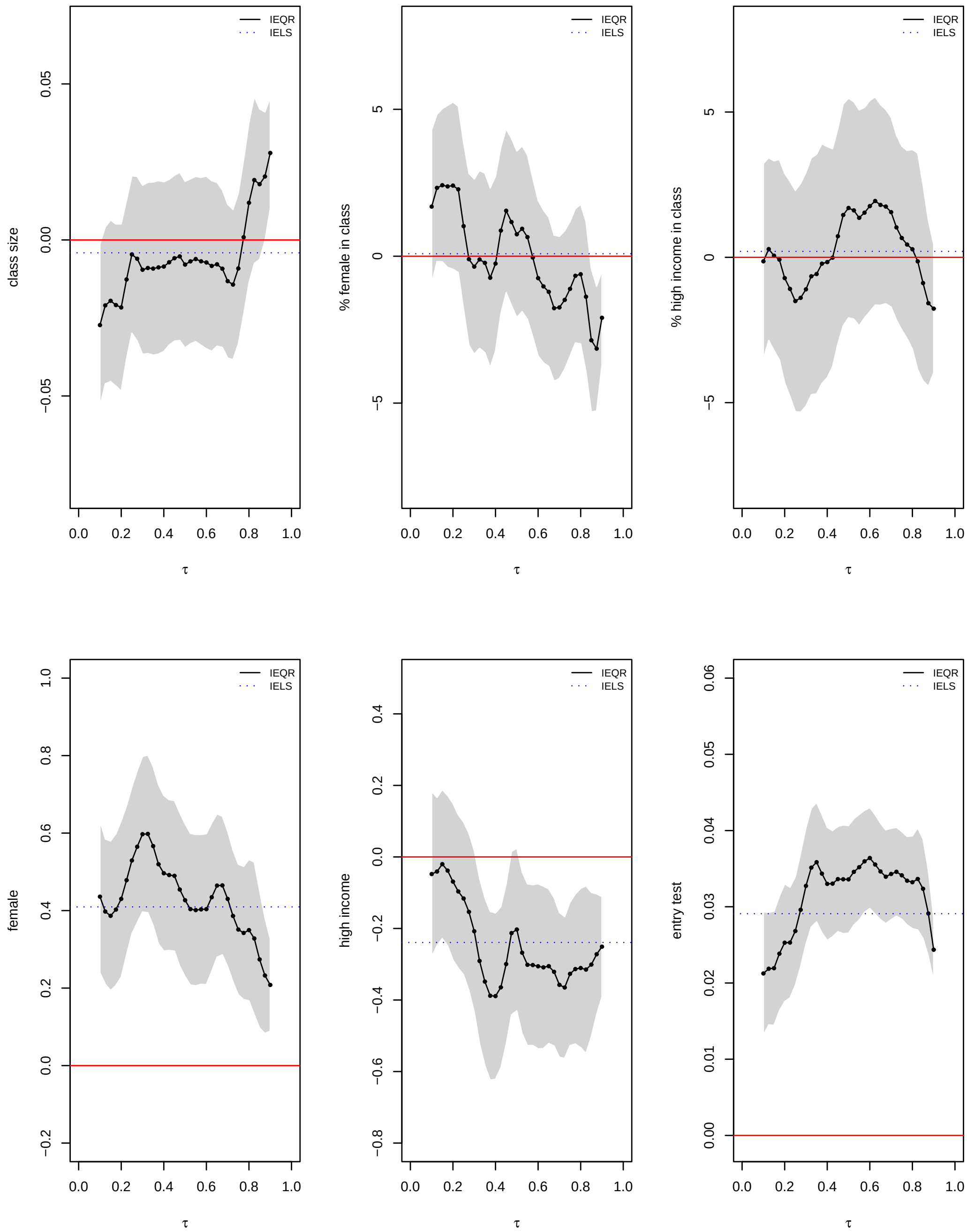

FIGURE 2. *Quantile regression covariate effects on educational attainment. The continuous dotted line shows quantile regression with interactive effects (QRIE) estimates and the dashed horizontal line shows the least squares version of the estimator (LSIE). These estimates were obtained from a sample of students in Management, and Finance.*

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 effective 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 teachers' allocation is not a concern in this data. Using Figure 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 5). 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 measure by percentage of female in class and percentage of high-income student in class do not seem to significantly affect performance. The sole exception is the effect of percentage of female in class at the 0.9 quantile.

Let us now consider the empirical evidence presented in Table 6. The table presents the estimated effects of the variables that may be of interest to policy makers: class size and class composition, measure by the percentages 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),

Method	Quantiles					Mean
	0.1	0.25	0.5	0.75	0.9	
			Class size effect			
Pooled methods	-0.005	-0.001	0.003	-0.006	-0.004	-0.002
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Instrumental Variable	-0.005	-0.001	-0.001	-0.007	-0.005	-0.003
	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)
Fixed Effects	-0.089	-0.088	-0.090	-0.088	-0.088	-0.093
	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)	(0.003)
Instrumental Variable	-0.014	-0.013	-0.001	0.006	0.005	-0.019
Fixed Effects	(0.002)	(0.003)	(0.006)	(0.001)	(0.001)	(0.007)
Interactive Effects	-0.031	-0.017	-0.008	-0.010	0.029	-0.007
	(0.015)	(0.017)	(0.016)	(0.014)	(0.011)	(0.011)
			% female in class			
Pooled methods	2.702	-0.426	0.732	-0.609	-1.302	-0.048
	(1.628)	(1.618)	(1.632)	(1.173)	(1.121)	(1.008)
Instrumental Variable	0.873	-1.616	-1.177	-1.313	-2.007	-0.797
	(1.276)	(1.225)	(1.211)	(0.922)	(0.897)	(1.049)
Fixed Effects	2.105	3.049	2.958	2.960	2.422	2.799
	(1.984)	(1.571)	(1.268)	(1.148)	(1.527)	(0.504)
Instrumental Variable	2.762	2.926	3.341	2.435	2.352	2.850
Fixed Effects	(0.551)	(0.871)	(1.940)	(0.284)	(0.226)	(0.703)
Interactive Effects	3.341	0.620	1.314	-0.479	-1.898	0.297
	(1.474)	(1.781)	(1.652)	(1.407)	(1.006)	(1.043)
			% high income in class			
Pooled methods	-3.289	-6.530	-6.931	-4.968	-4.498	-5.234
	(1.464)	(1.503)	(1.600)	(1.103)	(1.275)	(1.022)
Instrumental Variable	-1.725	-4.838	-6.388	-4.231	-4.003	-4.095
	(1.831)	(1.635)	(1.765)	(1.346)	(1.383)	(1.079)
Fixed Effects	3.798	4.187	3.626	3.701	3.420	4.273
	(1.947)	(1.582)	(1.342)	(1.384)	(1.765)	(0.657)
Instrumental Variable	4.897	5.111	5.039	4.289	4.181	5.504
Fixed Effects	(0.712)	(1.236)	(2.835)	(0.435)	(0.328)	(0.973)
Interactive Effects	0.774	-0.037	-0.181	-0.118	-1.587	0.147
	(1.658)	(2.024)	(2.014)	(1.626)	(1.210)	(1.239)

TABLE 6. *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.*

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 indicates that all students benefit from a class size reduction with stronger students and weak students 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 student, smaller classes have a negative influence on academic achievement.

The results of Table 6 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. The evidence 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. **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 τ_L percentile in the conditional distribution, to be at τ', 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,

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

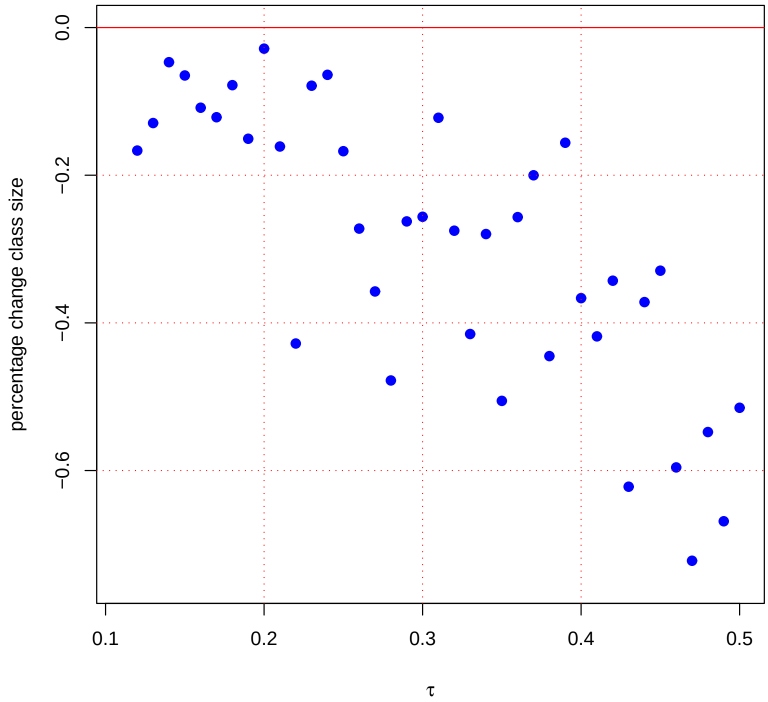

FIGURE 3. *Class size changes to prevent failure among the worst students.*

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 τ_j. An additional constraint is that the total number of students S are 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 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 $\tau = \{0.11, 0.12, \ldots, 0.5\}$. 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 5. 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.

8. Appendix

Proof of Proposition 1. Using assumption 1, we let $\boldsymbol{\omega} = \boldsymbol{F}\boldsymbol{\pi} + \boldsymbol{u}$. Then the estimator can be written as $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\boldsymbol{D}'\bar{\boldsymbol{P}}_{\bar{M}W}\boldsymbol{D})^{-1}\boldsymbol{D}'\bar{\boldsymbol{P}}_{\bar{M}W}(\boldsymbol{F}\boldsymbol{\lambda} + \boldsymbol{\omega})$ because $\bar{\boldsymbol{P}}_{\bar{M}W}\boldsymbol{X} = 0$. By Equation 2.7, we have $\bar{\boldsymbol{H}} = (\boldsymbol{G} + \bar{\boldsymbol{U}}^*)\bar{\boldsymbol{P}}^-$, where $\boldsymbol{G} = [\boldsymbol{X}, \boldsymbol{F}]$, and $\bar{\boldsymbol{U}}^* = [\boldsymbol{0}, (\bar{\boldsymbol{\xi}}' \otimes \iota_N)\boldsymbol{\Gamma}_0]$. The matrix $\bar{\boldsymbol{P}} = [\bar{\boldsymbol{P}}_1, -\bar{\boldsymbol{P}}_2]$, where $\bar{\boldsymbol{P}}_1 = (\boldsymbol{1}'_{k_2}, \boldsymbol{0}', \boldsymbol{0}', \boldsymbol{0}', \boldsymbol{0}')'$ and $\bar{\boldsymbol{P}}_2 = (\boldsymbol{0}', -\boldsymbol{\Gamma}'_0, \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \boldsymbol{\Gamma}_3)'$. Now by the employing the moment conditions on $\bar{\boldsymbol{U}}^*$ given in Assumptions 2 and 3, and after appropriately normalizing each term of the following expressions by NT^{-1} we have the following large N and large T convergence results:

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{D} \to \boldsymbol{D}\bar{\boldsymbol{M}}_q\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{D},$$

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{F} \to \boldsymbol{D}\bar{\boldsymbol{M}}_q\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{F},$$

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{\omega} \to \boldsymbol{D}\bar{\boldsymbol{M}}_q\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}_q\boldsymbol{\omega},$$

where $\bar{\boldsymbol{M}}_q = \boldsymbol{I} - \bar{\boldsymbol{Q}}(\bar{\boldsymbol{Q}}'\bar{\boldsymbol{Q}})^-\bar{\boldsymbol{Q}}$, and $\bar{\boldsymbol{Q}} = \boldsymbol{G}\bar{\boldsymbol{P}}^-$. Given the rank condition (Assumption 4), we have $\bar{\boldsymbol{M}}_q = \boldsymbol{M}_g := \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^-\boldsymbol{G}$. Therefore,

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{D} \to \boldsymbol{D}\boldsymbol{M}_g\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{M}_g\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{M}_g\boldsymbol{D},$$

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{F} \to \boldsymbol{0},$$

$$\boldsymbol{D}\bar{\boldsymbol{M}}\boldsymbol{W}(\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{W})^{-1}\boldsymbol{W}'\bar{\boldsymbol{M}}\boldsymbol{\omega} \to \boldsymbol{D}\boldsymbol{M}_g\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{M}_g\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{M}_g\boldsymbol{\omega},$$

Now apply Assumption 4 to see the last term goes to 0. By Assumption 1 the estimator is consistent. $\square$

Proof of Theorem 1. (Consistency) By proposition 1 in Chernozhukov and Hansen (2008), we have that $\sup_{\alpha\in\mathcal{A}} \|\hat{\boldsymbol{\vartheta}}(\alpha,\tau) - \boldsymbol{\vartheta}(\alpha,\tau)\| \to 0$ for $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{\gamma}')'$. This implies that $\sup_{\alpha\in\mathcal{A}} \|\hat{\boldsymbol{\gamma}}(\alpha,\tau) - \boldsymbol{\gamma}(\alpha,\tau)\| \to 0$, and that $\|\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}(\tau)\| \to 0$. Consider a small ball $\boldsymbol{\alpha}_n$ of radius r_n centered at $\alpha(\tau)$. Then for any $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}(\tau)$, we have that $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n,\tau) \to \boldsymbol{\beta}(\boldsymbol{\alpha}(\tau),\tau) = \boldsymbol{\beta}(\tau)$, $\hat{\boldsymbol{\lambda}}(\boldsymbol{\alpha}_n,\tau) \to \boldsymbol{\lambda}(\boldsymbol{\alpha}(\tau),\tau) = \boldsymbol{\lambda}(\tau)$, and $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_n,\tau) \to \boldsymbol{\gamma}(\boldsymbol{\alpha}(\tau),\tau) = \boldsymbol{\gamma}(\tau) = 0$. Hence $\hat{\boldsymbol{\vartheta}}(\boldsymbol{\alpha}_n,\tau) \to \boldsymbol{\vartheta}(\boldsymbol{\alpha}(\tau),\tau)$ for any $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}(\tau)$.

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

6, the solution of 3.5 is equivalent to the solution of minimizing,

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

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

$$\sup ||v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}) - \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0}))|| = o_p(1) \tag{8.1}$$

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

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

Taking expectation and expanding v under condition 5, we obtain

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

where $G(\cdot)$ is the conditional distribution of y. Clearly, $v(\hat{\boldsymbol{\delta}}_\alpha, \hat{\boldsymbol{\delta}}_\beta, \hat{\boldsymbol{\delta}}_\gamma, \hat{\boldsymbol{\delta}}_\lambda) \to 0$, 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,

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

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

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

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

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

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

$$\sup||v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0},\mathbf{0},\mathbf{0}) - \mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0},\mathbf{0},\mathbf{0}))|| = o_p(1) \tag{8.2}$$

Expanding as above we obtain

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

Notice that $v(\hat{\boldsymbol{\delta}}_\alpha, \hat{\boldsymbol{\delta}}_\beta, \hat{\boldsymbol{\delta}}_\gamma) \to 0$, 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 and $\boldsymbol{\delta}_\vartheta = (\boldsymbol{\delta}'_\beta, \boldsymbol{\delta}'_\gamma)'$, we write the last expression as,

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

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

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

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

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

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

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

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

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

□

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.

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, 3nd edn.

Bandiera, O., V. Larcinese, and I. Rasul (2010): "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," forthcoming, *Economic Journal*.

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.

Chudik, A., M. H. Pesaran, and E. Tosetti (2009): "Weak and Strong Cross Section Dependence and Estimation of Large Panels," mimeo.

Coleman, J. (1966): "Equality of Educational Opportunity," Report for US Department of Health, Education, and Welfare, Office of Education, Washington D.C.

Di 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.

Harding, M. C. (2007): "Structural Estimation of High Dimensional Factor Models," mimeo, MIT.

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.

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`.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

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

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.

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.