Estimating and Testing a Quantile Regression Model with Interactive Effects

Matthew Harding
Carlos Lamarche

DISCUSSION PAPER SERIES

I Z A

Estimating and Testing a Quantile Regression Model with Interactive Effects

Matthew Harding
Stanford University

Carlos Lamarche
*University of Oklahoma
and IZA*

ABSTRACT

Estimating and Testing a Quantile Regression Model with Interactive Effects[*]

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

Corresponding author:

Matthew Harding
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
USA
E-mail: mch@stanford.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. Our method differs from Harding and Lamarche (2009) and Galvao (2009) because it does not consider the case of unobserved heterogeneity represented by a classical individual effect λ_i. We provide conditions under which the slope parameter estimator is consistent and asymptotically Gaussian. Moreover, we investigate the finite sample performance of the proposed method in comparison to other candidate methods. Monte Carlo evidence shows that the finite sample performance of the proposed method is satisfactory in all the variants of the models, including specifications with $\boldsymbol{\lambda}$, $\boldsymbol{f}$ 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 a teacher's quality affects performance only if the student is motivated and receptive to instruction. We find that the proposed method gives different policy prescriptions relative to standard methods. While a reduction in class size does not impact mean and median student performance, it affects performance at the tails of the conditional distribution. Our finding suggests that this policy benefits weak students, but harms high achievers. Moreover, we find evidence that indicates that a change in the gender composition of a class impacts differently low- and high-performing students.

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

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

2. A Quantile Regression Approach

Consider the following model:

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

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

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 $\boldsymbol{\lambda}_i' \boldsymbol{f}_t$ 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,

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

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. It is known that an IV estimator for $\boldsymbol{\alpha}$ can be obtained in two steps from,

$$(2.4) \qquad \boldsymbol{y}(\boldsymbol{\alpha}) = \boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{F}\boldsymbol{\lambda} + \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{W}$ is a matrix of instruments (Chernozhukov and Hansen 2006). The latent term $\boldsymbol{F}$ can be approximated by cross-sectional averages of dependent and independent variables (Pesaran 2006).

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

ASSUMPTION 2. *The* $r \times 1$ *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* $\boldsymbol{v}_{it}'$ *for all* i,t,t'.

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

ASSUMPTION 4. *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.

We begin considering for simplicity the following conditional quantile functions:

$$(2.5) \quad Q_{Y_{it}}(\tau|\boldsymbol{d}_{it}, \boldsymbol{x}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t) = \boldsymbol{\alpha}'\boldsymbol{d}_{it} + \boldsymbol{\beta}'\boldsymbol{x}_{it} + \boldsymbol{\lambda}_i'\boldsymbol{f}_t + G_u(\tau)^{-1},$$

$$(2.6) \quad Q_{D_{it}}(\tau|\boldsymbol{w}_{it}, \boldsymbol{x}_{it}, \boldsymbol{\lambda}_i, \boldsymbol{f}_t) = \boldsymbol{\Pi}_1'\boldsymbol{w}_{it} + \boldsymbol{\Pi}_2'\boldsymbol{x}_{it} + \boldsymbol{\Pi}_{3i}'\boldsymbol{f}_t + \boldsymbol{\Pi}_4\boldsymbol{\lambda}_i'\boldsymbol{f}_t + \boldsymbol{\Pi}_5'\boldsymbol{\lambda}_i + \kappa G_u(\tau)^{-1},$$

where τ is a quantile in the interval $(0,1)$, κ is a parameter that might be interpreted as measuring the conditional correlation between Y and D, and G denotes the distribution function of the *iid* error term u. This model (2.5)-(2.6) is a simple quantile regression version of model (2.1)-(2.2). A natural generalization of the model is defining the quantile function (2.6) for $G_v(\tau')^{-1}$. To estimate this model, we can accommodate the instrumental variable approach proposed in Chernozhukov and Hansen (2005) to panel data, integrating out the quantile τ' as in Ma and Koenker (2006). If we augment the design matrix with cross-sectional averages, we have that the approximation for the factors $\boldsymbol{f}$'s depend on the quantiles τ and τ'. By integrating out the quantile τ', we may define the factors in terms of τ. Chernozhukov and Hansen (2005) illustrate that their approach is always applicable to a class of triangular models.

The model can be easily generalized to the standard location-scale shift model and other more general panel data models with heterogeneous effects. As it will be clear in Section 5, we will use the general version of these equations as a model for educational achievement. Following the literature (see, e.g., Ma and Koenker 2006, Hanushek et al. 2003, De Giorgi, Pellizzari and Woolston 2009, Bandiera, Larcinese, and Rasul 2010), the response variable y is educational attainment and is influenced by class size and peer effects $\boldsymbol{d}$, and individual, family and school characteristics $\boldsymbol{x}$. The last term u may represent idiosyncratic shocks to achievement which force the student to switch class.

A central concern in the estimation of the distributional effects of class size is unobserved heterogeneity. While most of the models estimated in the literature assume the classical additive separable structure on unobserved heterogeneity $\lambda + f$, we will estimate a more general specification allowing for interactive effects. The variable $\boldsymbol{\lambda}$ captures a student's unobserved ability to absorb knowledge when listening to lectures, effort and motivation, and the variable $\boldsymbol{f}$ includes teachers' quality and other class-invariant unobserved effects. The education production function (2.5) incorporates unobserved heterogeneity, while allowing for the possibility that teachers' quality affects performance *only if* the student is motivated and receptive to instruction.

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 (2.6) into (2.5) and summing over the cross-sectional

dimension of the model, we obtain,

$$(2.7) \qquad \bar{z}_t(\tau) = C_1 \bar{w}_t + C_2(\tau) \bar{x}_t + (\bar{C}_3 + C_4 \bar{\lambda}') f_t + C_5 \bar{\lambda},$$

where $\bar{z}_t(\tau)$ is the cross-sectional average of $z_{it}(\tau) = (Q_{Y_{it}}(\tau|\bullet), Q_{D_{it}}(\tau|\bullet)')'$, and $C_1 = (\alpha' \Pi_1', \Pi_1')'$, $C_2(\tau) = ((\alpha' \Pi_2' + \beta(\tau)')', \Pi_2')'$, $\bar{C}_3 = N^{-1} \sum_{i=1}^N ((\Pi_{3i} \alpha)', \Pi_{3i}')'$, $C_4 = (\alpha' \Pi_4 + 1, \Pi_4)'$, $\bar{\lambda} = N^{-1} \sum_{i=1}^N \lambda_i$ and $C_5 = (\alpha' \Pi_5, \Pi_5)'$.

ASSUMPTION 5. *The matrix $\bar{C}_3 + C_4 \bar{\lambda}'$ converges to a limiting matrix with rank $k_1 + 1 < r$.*

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

$$(2.8) \qquad C_{it}(\tau, \alpha, \beta, \delta, \gamma) = \rho_\tau \left(y_{it} - d_{it}' \alpha - x_{it}' \beta - \hat{f}_t'(\tau) \delta - \hat{\Phi}_{it}'(\tau) \gamma \right).$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the standard quantile loss function (see, e.g., Koenker 2005). The quantile regression check function includes two additional terms that deserve our attention. Note that equation (2.7) suggests that unknown factors can be approximated by a term $f_t(\tau) = \Psi(\tau; \bar{z}_t, \bar{w}_t, \bar{x}_t, 1)$, which is a known parametric function of cross-sectional averages of the endogenous and exogenous variables. A practical formulation would be to use a vector $\hat{f}_t(\tau)$ that includes an intercept and the cross-sectional variables $\hat{z}_t(\tau)$, $\bar{w}_t$, and $\bar{x}_t$. An alternative formulation for $\bar{z}_t(\tau)$ is to use cross-sectional averages of $z_{it} = (y_{it}, d_{it}')'$ as suggested in Pesaran (2006). If cross-sectional averages are used instead of the conditional quantiles Q_D's, the vector d_{it} in (2.8) is replaced by $\tilde{d}_{it} = (d_{it}', \bar{d}_t')'$. Notice that we may instrument $\tilde{d}_{it}$ by the vector of instruments $\tilde{w}_{it} = (w_{it}', \bar{w}_t')'$. In this case, $f_t(\tau)$ is defined as $\Psi(\tau; \bar{y}_t, \bar{x}_t, 1)$.

The second term $\Phi_{it}(\tau) = \Phi(\tau; w_{it}, x_{it}, f_t, \lambda_i)$ is a vector of transformations of instruments as introduced by Chernozhukov and Hansen (2006). In practice, it is possible to estimate Φ by a least squares projection of the endogenous variables d_{it} on the instruments w_{it}, the exogenous variables x_{it}, and a vector of individual and time effects. Alternatively, the interactive effects might be obtained by the least squares approaches proposed by Pesaran (2006) and Bai (2009). We consider the case of $\dim(\alpha) = \dim(\gamma)$, although the vector Φ may include more elements than the vector d_{it}.

The procedure is similar in spirit to Chernozhukov and Hansen (2006) applied to panel models as in Harding and Lamarche (2009) and Galvao (2009). First, we minimize the objective function

above for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$ as functions of τ and $\boldsymbol{\alpha}$,

$$(2.9) \qquad \{\hat{\boldsymbol{\beta}}(\tau,\boldsymbol{\alpha}), \hat{\boldsymbol{\delta}}(\tau,\boldsymbol{\alpha}), \hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})\} = \underset{\beta,\gamma,\delta \in \mathcal{B}\times\mathcal{G}\times\mathcal{F}}{\operatorname{argmin}} \sum_{t=1}^{T}\sum_{i=1}^{N} C_{it}(\tau,\boldsymbol{\alpha};\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\gamma}).$$

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

$$(2.10) \qquad \hat{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})'\hat{\boldsymbol{A}}(\tau)\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha}) \right\},$$

for a positive definite matrix $\boldsymbol{A}$.[1] Then, the quantile regression estimator for a model with interactive effects (QRIE) is defined as,

$$(2.11) \qquad \hat{\boldsymbol{\theta}}(\tau) \equiv \left(\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\tau), \hat{\boldsymbol{\delta}}(\tau)\right) = \left(\hat{\boldsymbol{\alpha}}(\tau), \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}(\tau),\tau)), \hat{\boldsymbol{\delta}}(\hat{\boldsymbol{\alpha}}(\tau),\tau))\right).$$

It is straightforward to accommodate our estimator to the case of individual location shifts considered in Koenker (2004).[2] We define,

$$(2.12) \qquad C_{it}(\tau,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{\gamma},\mu_i) = \rho_\tau\left(y_{it} - \boldsymbol{d}_{it}'\boldsymbol{\alpha} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \mu_i - \hat{\boldsymbol{f}}_t'(\tau)\boldsymbol{\delta} - \hat{\boldsymbol{\Phi}}_{it}'(\tau)\boldsymbol{\gamma}\right).$$

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

$$(2.13) \qquad \{\tilde{\boldsymbol{\beta}}(\tau,\boldsymbol{\alpha}), \tilde{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha}), \tilde{\boldsymbol{\mu}}(\tau,\boldsymbol{\alpha}), \tilde{\boldsymbol{\delta}}(\tau,\boldsymbol{\alpha})\} = \underset{\beta,\gamma,\delta,\mu \in \mathcal{B}\times\mathcal{G}\times\mathcal{F}\times\Lambda}{\operatorname{argmin}} \sum_{t=1}^{T}\sum_{i=1}^{N} C_{it}(\tau,\boldsymbol{\alpha};\boldsymbol{\beta},\boldsymbol{\mu},\boldsymbol{\delta},\boldsymbol{\gamma}).$$

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

$$(2.14) \qquad \tilde{\boldsymbol{\alpha}}(\tau) = \underset{\boldsymbol{\alpha}\in\mathcal{A}}{\operatorname{argmin}} \left\{ \tilde{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})'\tilde{\boldsymbol{A}}(\tau)\tilde{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha}) \right\},$$

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

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

---

[1]Although the exact form of $\boldsymbol{A}(\tau)$ is not important, it is often convenient to consider the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\alpha})$. This leads to a chi-square property for $\hat{\boldsymbol{\gamma}}(\tau)'\boldsymbol{A}(\tau)\hat{\boldsymbol{\gamma}}(\tau)$ (Chernozhukov and Hansen 2006). The requirement is that $\boldsymbol{A}(\tau)$ be any uniformly positive definite matrix, and for practical convenience, can be $\boldsymbol{A}(\tau) = \boldsymbol{I}$ or $\hat{\boldsymbol{A}}(\tau) = (NT)^{-1}\sum_i\sum_t \boldsymbol{\Phi}_{it}(\tau)\boldsymbol{\Phi}_{it}(\tau)'$.

[2]In the spirit of Bai (2009, section 8), we accommodate our estimator to models with both additive individual effects and interactive individual effects. Although we do not explicitly introduce a model with additive individual effects, it is possible to obtain an additive individual effect and an interactive effects term in equation (2.5), by noting that $\boldsymbol{\lambda}_i'\boldsymbol{f}_t = \lambda_{i1} + \sum_{k=2}^{r}\lambda_{ik}f_{tk}$, if $f_{t1} = 1$ for all t.

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

Example 1. Consider the case that $f_t = r = 1$ for all t. For simplicity, we let $\mathbf{\Pi}_{3i} = \mathbf{0}$ in equation (2.6) and $\tilde{C}_4 = C_4 + C_5$. From equation (2.7), we have that, $\tilde{C}_4 \lambda_i = z_{it}(\tau) - C_1 w_{it} - C_2(\tau) x_{it}$, or, similarly, by taking averages over the time-series dimension of the model, $\lambda_i = \mathbf{\Psi}(\tau)' \bar{s}_i$, where $s_{it} = (z_{it}(\tau)', w_{it}', x_{it}')'$, $\bar{s}_i$ denotes the sample average of s, and $\mathbf{\Psi}(\tau)$ is a vector of coefficients that includes $\tilde{C}_4$, C_1, and C_2. Therefore, in the one-factor model with fixed effects λ_i's, we can augment the model by a vector of observables $\bar{s}_i$. In the linear least squares case, it is known that the classical fixed effects estimator is numerically equivalent to the estimator obtained by augmenting the model by a vector of observables (Chamberlain 1982, Mundlak 1978).

Example 2. In the simplest conceivable case of $\boldsymbol{f}_t$ not depending on the quantile τ and u_{it} stochastically independent on $\boldsymbol{v}_{it}$, one can interpret that the method considers proxing $\boldsymbol{f}_t$ with $\bar{s}_t = (\bar{z}_t', \bar{x}_t')'$, implicitly assuming that $\boldsymbol{f}_t = \mathbf{\Pi}' \bar{s}_t + \boldsymbol{\epsilon}_t$. To address the possibility that the loadings $\boldsymbol{\lambda}_i$'s are correlated with the independent variables, Bai (2009) notices that one can similarly write $\boldsymbol{\lambda}_i = \mathbf{\Psi}' \bar{s}_i + \boldsymbol{\eta}_i$, where $\bar{s}_i = (\bar{d}_i, \bar{x}_i')'$. The errors $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_i$ are assumed to be distributed as F_ϵ and F_η, and are independent of the covariates. Notice that, $\boldsymbol{\lambda}_i' \boldsymbol{f}_t = \bar{s}_i' \mathbf{\Theta} \bar{s}_t + \boldsymbol{\theta}_i' \bar{s}_t + \bar{s}_i' \boldsymbol{\theta}_t + \boldsymbol{\eta}_i' \boldsymbol{\epsilon}_t$, where $\mathbf{\Theta} = \mathbf{\Psi}' \mathbf{\Pi}$ is a $p \times p$ matrix, $\boldsymbol{\theta}_i = (\boldsymbol{\eta}_i' \mathbf{\Pi})'$ is a $p \times 1$ vector, and $\boldsymbol{\theta}_t = (\mathbf{\Psi}' \boldsymbol{\epsilon}_t)'$ is a $p \times 1$ vector. Under the assumption that $\boldsymbol{\eta}_i' \boldsymbol{\epsilon}_t$ and the right hand side variables are independent, we can augment the model and consistently estimate the effects of interest using an approach similar to the one presented below (see, e.g., Bai 2009, Abrevaya and Dahl 2008).

Example 3. The one-factor model case suggests that the coefficient $\boldsymbol{\delta}(\tau)$ in equation (2.8) may be interpreted as a reduced form coefficient. From equation (2.7), we can write,

$$(2.15) \qquad f_t = \mathbf{\Gamma}_0' \bar{z}_t(\tau) - \mathbf{\Gamma}_0' C_1 \bar{w}_t - \mathbf{\Gamma}_0' C_2(\tau) \bar{x}_t - \mathbf{\Gamma}_0' C_5 \bar{\lambda},$$

where as before $\mathbf{\Gamma}_0 = \tilde{C}_3 (\tilde{C}_3' \tilde{C}_3)^{-1}$ and $\tilde{C}_3 = \bar{C}_3 + C_4 \bar{\lambda}'$. Multiplying (2.15) by the one-dimensional loading λ_i, we have that,

$$(2.16) \qquad \lambda_i f_t = \lambda_i \mathbf{\Gamma}_0' \bar{z}_t(\tau) - \lambda_i \mathbf{\Gamma}_0' C_1 \bar{w}_t - \lambda_i \mathbf{\Gamma}_0' C_2(\tau) \bar{x}_t - \lambda_i \mathbf{\Gamma}_0' C_5 \bar{\lambda}$$

$$(2.17) \qquad \qquad = \boldsymbol{\delta}_0' \bar{z}_t(\tau) + \boldsymbol{\delta}_1' \bar{w}_t + \boldsymbol{\delta}_2(\tau)' \bar{x}_t + \boldsymbol{\delta}_5' \bar{\lambda},$$

where $\boldsymbol{\delta}_0 = \mathbf{\Gamma}_0 \lambda_i$, $\boldsymbol{\delta}_1 = -\boldsymbol{\delta}_0 C_1$, $\boldsymbol{\delta}_2(\tau) = -\boldsymbol{\delta}_0 C_2(\tau)$, and $\boldsymbol{\delta}_5 = -\boldsymbol{\delta}_0 C_5$. Note that the reduced form parameter $\boldsymbol{\delta}_0$ is then,

$$\boldsymbol{\delta}_0 = \mathbf{\Gamma}_0 \lambda_i = \frac{1}{N \tilde{C}_3' \tilde{C}_3} \sum_{i=1}^N \left(\begin{array}{c} \boldsymbol{\alpha}' \mathbf{\Pi}_{3i} \lambda_i + (\boldsymbol{\alpha}' \mathbf{\Pi}_4 + 1) \lambda_i^2 \\ \mathbf{\Pi}_{3i} \lambda_i + \mathbf{\Pi}_4 \lambda_i^2 \end{array} \right) = \frac{1}{\tilde{C}_3' \tilde{C}_3} \left(\begin{array}{c} \boldsymbol{\alpha}' \bar{\lambda}_{\Pi_3} + (\boldsymbol{\alpha}' \mathbf{\Pi}_4 + 1) \bar{\lambda}^2 \\ \bar{\lambda}_{\Pi_3} + \mathbf{\Pi}_4 \bar{\lambda}^2 \end{array} \right).$$

where $\bar{\lambda}_{\Pi_3} = N^{-1} \sum_i \Pi_{3i} \lambda_i$ and $\bar{\lambda}^2 = N^{-1} \sum_i \lambda_i^2$. Using equation (2.17) and letting $\boldsymbol{f}(\tau) = (\bar{\boldsymbol{z}}_t(\tau)', \bar{\boldsymbol{w}}_t', \bar{\boldsymbol{x}}_t', \boldsymbol{\iota})'$ and $\boldsymbol{\delta}(\tau) = (\boldsymbol{\delta}_0', \boldsymbol{\delta}_1', \boldsymbol{\delta}_2(\tau)', \boldsymbol{\delta}_5')'$, we have that $\lambda_i f_t$ equals $f_t(\tau)\delta(\tau)$.

Remark 1. The conditional quantile function (2.5) 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 also be simply estimated by the method developed by Koenker. A model with endogenous covariates can be estimated by the method proposed in Harding and Lamarche (2009).

Remark 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 in 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). 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 $\|\hat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\alpha})\|_{\hat{\boldsymbol{A}}(\tau)}^2$.

3. Additional Assumptions and Basic Inference

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

ASSUMPTION 6. *The variables y_{it} are independent across subjects i with conditional distribution functions G_{it}, differentiable conditional densities, $0 < g_{it} < \infty$, with bounded derivatives g_{it}' at the conditional quantiles $\xi_{it}(\tau)$ for $i = 1, ..., N$ and $t = 1, ..., T$.*

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,*

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

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

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

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

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

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

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 6 ensures a well-defined asymptotic behavior of the quantile regression estimator. Condition 4 implies the standard conditions on the instruments in the exactly identified case. This case can be easily relaxed but we impose it by convenience. The independence on the y_{it}'s is assumed in Koenker (2004) and conditions 7 and 8 are assumed in Chernozhukov and Hansen (2006). In condition 9, the existence of the limiting form of the positive definite matrices is used to invoke the Lindeberg-Feller Central Limit Theorem. The asymptotic covariance matrices have a representation similar to the matrices found in Chernozhukov and Hansen (2008, proposition 2) and Galvao (2009, theorem 2). Condition 10 is important both for the Lindeberg condition and for ensuring the finite dimensional convergence of the objective function.

PROPOSITION 1. *Under regularity conditions A1-A10, the quantile regression estimator (with interactive effects) defined in (2.11),* $(\hat{\boldsymbol{\alpha}}(\tau)', \hat{\boldsymbol{\beta}}(\tau)')'$, *is consistent and asymptotically normally distributed with mean* $(\boldsymbol{\alpha}(\tau)', \boldsymbol{\beta}(\tau)')'$ *and covariance matrix* $\boldsymbol{J}'(\tau)\boldsymbol{S}(\tau)\boldsymbol{J}(\tau)$.

The estimation of $\boldsymbol{S}$ and $\boldsymbol{J}$ depends on conditional densities and factors, and consequently, they cannot be directly obtain from the sample counterparts. We propose to use a kernel estimator to estimate the densities (Koenker 2005, Kato and Galvao 2010) and cross-sectional averages to estimate the factors (Pesaran 2006). The matrix of factors $\boldsymbol{F}$ of dimension $NT \times r$ is replaced by a

matrix $\hat{\boldsymbol{F}}$ of dimension $NT \times 1 + k_1 + k_2$. Let h denote the bandwidth and $\hat{g} = h^{-1}K(u/h)$ where $K(\cdot)$ is a continuous, bounded kernel function and $\hat{u}_{it}(\tau) = y_{it} - \boldsymbol{d}'_{it}\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau) - \hat{\boldsymbol{f}}'_t\hat{\boldsymbol{\delta}}(\tau)$ is a residual associated with u. The bandwidth $h \to 0$ as N and $T \to \infty$. We define the sample analog of $\boldsymbol{S}$ as, $\hat{\boldsymbol{S}}_{NT}(\hat{\boldsymbol{\theta}}) = (\tau(1-\tau))(NT)^{-1}\sum_i\sum_t(\tilde{m}_{it}(\hat{\boldsymbol{\theta}})\tilde{\boldsymbol{x}}_{it})(\tilde{m}_{it}(\hat{\boldsymbol{\theta}})\tilde{\boldsymbol{x}}_{it})'$.

Remark 3. The uniform convergence of the asymptotic covariance matrix can be verified elementwise using Corollary 3.1 of Newey (1991). By Assumption 7, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, a compact set. Moreover, Assumptions 1-5 satisfy the conditions of Lemma 1 in Pesaran (2006), then the vector of observable proxies $\hat{\boldsymbol{f}}_t$ is consistent for $\boldsymbol{f}_t$. The consistency of $\hat{\boldsymbol{S}}_{NT}(\hat{\boldsymbol{\theta}})$ also requires the consistency of $\hat{\boldsymbol{\theta}}$, which is shown in the first part of Proposition 1. Under conditions similar to ours, provided that $\hat{\boldsymbol{\theta}}$ is a consistent estimator as shown in Proposition 1, the literature shows the consistency of the kernel estimator. Additionally, by the Law of Large Numbers and conditions 1 and 10, we have that a re-centered $\hat{\boldsymbol{S}}_{NT}(\boldsymbol{\theta})$ converges in probability to zero for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The present conditions satisfy the conditions of Corollary 3.1 of Newey (1991) and therefore, the uniform convergence of $\hat{\boldsymbol{S}}_{NT}(\boldsymbol{\theta})$ is established. Similar arguments are applied to $\hat{\boldsymbol{J}}_{NT}(\boldsymbol{\theta})$ under continuity of matrix inversion. By the Slutsky theorem, $\hat{\boldsymbol{J}}'\hat{\boldsymbol{S}}\hat{\boldsymbol{J}}$ consistently estimates the asymptotic covariance matrix $\boldsymbol{J}'\boldsymbol{S}\boldsymbol{J}$.

Inference can be considered by accommodating standard tests. Wald-type statistics (Koenker and Bassett 1982) can be considered for basic linear hypothesis on a vector $\boldsymbol{\theta}$. More general hypothesis including evaluating the vector over a range of quantiles could be accommodated by considering the Kolmogorov-Smirnov statistics (Koenker and Xiao 2002). We consider a general linear hypothesis on the vector $\boldsymbol{\theta}$ of the form $H_0 : \boldsymbol{R}\boldsymbol{\theta} = \boldsymbol{r}$, where $\boldsymbol{R}$ is a matrix that depends on the type of restrictions imposed. The test statistics is defined as,

$$(3.1) \qquad \mathcal{T}_{NT} = NT(\boldsymbol{R}\hat{\boldsymbol{\theta}} - \boldsymbol{r})'\left[\boldsymbol{R}\hat{\boldsymbol{V}}\boldsymbol{R}'\right]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\theta}} - \boldsymbol{r}),$$

is asymptotically distributed as χ_q^2 under H_0, where q is the rank of the matrix $\boldsymbol{R}$ and $\hat{\boldsymbol{V}}$ is the estimated covariance matrix, $\hat{\boldsymbol{J}}'\hat{\boldsymbol{S}}\hat{\boldsymbol{J}}$. This framework allows us to evaluate several linear hypotheses, including the significance of differences across coefficient estimates. Moreover, it is possible to apply this framework to obtain Hausman (1978) type tests for exogeneity (Chernozhukov and Hansen 2006) and interactive effects versus additive effects (Bai 2009).

4. 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 generate the dependent variable considering a

design similar to Bai (2009) and Pesaran (2006):

$$(4.1) \qquad y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 x_{it} + \lambda_{1i} f_{1t} + \lambda_{2i} f_{2t} + (1 + h d_{it}) u_{it}$$

$$(4.2) \qquad d_{it} = \pi_0 + \pi_1 w_{it} + \pi_2 x_{it} + \pi_3 f_{1t} + \pi_3 f_{2t} + \pi_4 \lambda_{1i} f_{1t} + \pi_4 \lambda_{2i} f_{2t} + v_{it}$$

$$(4.3) \qquad w_{it} = \theta_0 \mu_t + \theta_1 \epsilon_{it},$$

$$(4.4) \qquad f_{jt} = \rho_f f_{jt-1} + \eta_{jt}$$

$$(4.5) \qquad \eta_{jt} = \rho_\eta \eta_{jt-1} + e_{jt}$$

for $j = \{1, 2\}, \ldots t = -49, \ldots 0, \ldots T$ in the last two equations. We adopted $t_{-50} = -49$ following the design considered in Pesaran (2006). The error terms are $(u_{it}, v_{it})' \sim \mathcal{N}(0, \boldsymbol{\Omega})$ and x_{it}, μ_t, ϵ_{it}, e_{it} are Gaussian independent random variables. The loading $\lambda_{i1} \sim \mathcal{N}(1, 0.2)$ and λ_{i2} is distributed either as $\mathcal{N}(1, 0.2)$ or t-student distribution with two degrees of freedom. The parameters are assumed to be: $\beta_0 = \pi_0 = 0$, $\beta_1 = \beta_2 = \pi_1 = \pi_2 = \pi_3 = \theta_0 = \theta_1 = 1$, $\rho_f = 0.90$, $\rho_\eta = 0.25$, and $\Omega_{11} = \Omega_{22} = 1$. We consider three designs for the location shift model $h = 0$:

Design 1: The endogenous variable d is not correlated with the 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 (4.1) and (4.2) are not correlated. We assume $\pi_4 = 0.1$ and $\Omega_{12} = \Omega_{21} = 0$.

Design 3: The error terms in equations (4.1) and (4.2) 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.

Moreover, we expand the analysis considering these designs for the case of location-scale shift models. In all the variants of the experiments, we assume that $h = 0.1$ in equation (4.1).

Table 4.1 presents the bias and root mean square error (RMSE) for the slope parameter. While the first four columns of the table present results for the case that λ_1 and λ_2 are distributed as Gaussian random variables ($\mathcal{N}(1, 0.2)$), the last four columns of the table present results for the case of λ_2 distributed as t-student distribution with two degrees of freedom (t_2). We present results based on $N = \{100, 250\}$ and $T = \{4, 8\}$. It is important to note that even though the asymptotic results were derived under the assumption of large N and large T, the finite sample performance of the proposed estimators is excellent even for a very small number of time periods. The tables show results obtained from: the classical quantile regression estimator (QR), the fixed effects version of the estimator introduced by Koenker (2004) labelled QRFE, the instrumental variable method with

Statistic	Sample		Quantile Regression Panel Data Methods							
	Size		Gaussian distribution				t_2 distribution			
			QR	QRFE	QRIVFE	QRIE	QR	QRFE	QRIVFE	QRIE
	N	T				Design 1				
Bias	100	4	0.2695	0.3038	-0.2319	0.0092	0.1542	0.1783	-0.0664	0.0098
RMSE	100	4	0.3352	0.3700	1.6578	0.1706	0.2337	0.2567	0.4506	0.2324
Bias	250	4	0.2766	0.3141	-0.1863	-0.0010	0.1343	0.1525	-0.2122	0.0021
RMSE	250	4	0.3345	0.3751	4.6359	0.1048	0.2033	0.2300	1.9292	0.1626
Bias	100	8	0.4106	0.4230	-0.1291	-0.0138	0.2009	0.2154	-0.0616	-0.0075
RMSE	100	8	0.4498	0.4609	0.8426	0.1404	0.2703	0.2753	0.6111	0.2142
Bias	250	8	0.3864	0.4002	-0.1341	-0.0004	0.2126	0.2244	-0.2297	0.0057
RMSE	250	8	0.4251	0.4382	1.1526	0.0839	0.2699	0.2792	0.8133	0.1110
	N	T				Design 2				
Bias	100	4	0.3311	0.3169	-0.2621	0.0085	0.4254	0.2372	-0.0784	0.0141
RMSE	100	4	0.3783	0.3746	2.1627	0.1721	0.5365	0.2937	0.5127	0.2435
Bias	250	4	0.3417	0.3264	-0.0582	-0.0024	0.4620	0.2119	-0.0897	0.0092
RMSE	250	4	0.3866	0.3813	2.7423	0.1088	0.6094	0.2702	0.8594	0.1622
Bias	100	8	0.4480	0.4304	-0.1850	-0.0136	0.4398	0.2792	-0.0949	-0.0064
RMSE	100	8	0.4781	0.4624	1.3522	0.1382	0.5145	0.3215	0.5524	0.2101
Bias	250	8	0.4300	0.4115	-0.4482	0.0016	0.4421	0.2925	-0.3328	0.0029
RMSE	250	8	0.4583	0.4433	5.4814	0.0853	0.5153	0.3272	1.4542	0.1198
	N	T				Design 3				
Bias	100	4	0.4710	0.4410	-0.2639	0.0149	0.5513	0.3502	-0.0906	0.0036
RMSE	100	4	0.4962	0.4711	2.1051	0.1806	0.6415	0.3821	0.5747	0.2676
Bias	250	4	0.4714	0.4397	-0.0582	0.0022	0.5937	0.3330	-0.1188	0.0041
RMSE	250	4	0.4971	0.4712	2.7774	0.1022	0.7166	0.3678	0.9933	0.1731
Bias	100	8	0.5459	0.5229	-0.1736	0.0092	0.5387	0.3785	-0.0856	-0.0067
RMSE	100	8	0.5636	0.5422	1.3123	0.1331	0.6002	0.4061	0.5913	0.1955
Bias	250	8	0.5344	0.5069	-0.4783	0.0061	0.5437	0.3912	-0.3240	0.0077
RMSE	250	8	0.5508	0.5252	5.8590	0.0889	0.6019	0.4130	1.4801	0.1247

TABLE 4.1. *Small sample performance of a class of panel data estimators. The evidence is based on 400 randomly generated samples.*

fixed effects (QRIVFE) introduced by Harding and Lamarche (2009), and the quantile regression estimator for a model with interactive effects (QRIE).

Table 4.1 provides evidence of the biases present in the application of quantile 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 the pooled QR method is biased and has poor MSE properties, it is surprising that the application of QRFE estimation to a situation with interactive effects produces large

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

TABLE 4.2. *Small sample performance of quantile regression estimators in the location-scale shift model.*

biases. 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 QRIE estimator does not deteriorate in the presence of outliers, i.e. if the loading λ_2 follows a t_2 distribution. The quantile estimator QRIE continues to perform well with only a minor increase in MSE. The bottom part of Table 4.1 confirms that the above interpretation of the relative performance of the different estimators

Statistic	Sample		0.25 Quantile				0.75 Quantile		
	N	T	QRFE	QRIVFE	QRIE	QRIIE	QRFE	QRIVFE	QRIE	QRIIE
Design 4: Model with individual effects										
Bias	100	4	0.0230	0.0815	0.0529	0.0358	-0.0134	-0.0824	-0.0221	-0.0262
RMSE	100	4	0.0481	1.2867	0.1545	0.1440	0.0493	1.0241	0.1445	0.1482
Bias	100	6	0.0183	0.0027	0.0266	0.0283	-0.0133	-0.0185	-0.0178	-0.0365
RMSE	100	6	0.0355	0.2244	0.1265	0.1105	0.0334	0.2244	0.1537	0.1143
Bias	100	12	0.0076	-0.0464	0.0307	0.0221	-0.0086	-0.1739	-0.0283	-0.0257
RMSE	100	12	0.0195	1.1278	0.0994	0.0955	0.0210	2.2456	0.1120	0.0980
Bias	100	20	0.0065	0.0564	0.0312	0.0212	-0.0067	-0.0543	-0.0334	-0.0254
RMSE	100	20	0.0134	0.3401	0.0986	0.0763	0.0134	0.2640	0.0911	0.0750
Design 5: Model with individual effects and interactive effects										
Bias	100	4	0.3720	0.0419	0.0537	0.0467	0.3051	-0.3001	-0.0006	-0.0196
RMSE	100	4	0.4305	2.7468	0.2222	0.1935	0.3692	3.3622	0.2120	0.2002
Bias	100	6	0.3826	-0.1634	0.0338	0.0355	0.3227	-0.1244	-0.0274	-0.0443
RMSE	100	6	0.4258	0.9781	0.1990	0.1321	0.3707	0.9931	0.1999	0.1532
Bias	100	12	0.5082	-1.6556	0.0359	0.0271	0.4580	-2.7600	-0.0288	-0.0296
RMSE	100	12	0.5318	24.9890	0.1342	0.1167	0.4836	40.8708	0.1577	0.1303
Bias	100	20	0.5972	-0.0713	0.0360	0.0315	0.5328	-0.2495	-0.0399	-0.0335
RMSE	100	20	0.6133	0.8050	0.1125	0.1010	0.5490	1.0512	0.1147	0.1078
Design 6: Model with endogenous covariates and individual and interactive effects										
Bias	100	4	0.4861	-0.0339	0.0253	0.0320	0.4075	-0.2240	-0.0130	-0.0049
RMSE	100	4	0.5228	2.3460	0.2099	0.2024	0.4484	4.1126	0.2124	0.2087
Bias	100	6	0.4918	-0.1659	0.0218	0.0161	0.4253	-0.1313	-0.0217	-0.0311
RMSE	100	6	0.5194	0.9095	0.1877	0.1394	0.4559	0.9775	0.2126	0.1554
Bias	100	12	0.6027	-1.7410	0.0129	0.0158	0.5263	-1.6913	-0.0106	-0.0216
RMSE	100	12	0.6174	26.3003	0.1437	0.1234	0.5434	24.0678	0.1625	0.1286
Bias	100	20	0.6655	-0.0660	0.0220	0.0194	0.5843	-0.2332	-0.0374	-0.0212
RMSE	100	20	0.6771	0.7962	0.1119	0.0974	0.5956	1.0021	0.1159	0.1088

TABLE 4.3. *Small sample performance of quantile regression estimators in the location-scale shift model.*

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.

Table 4.2 presents the bias and root mean square error in the Gaussian case when β_1 in equation 4.1 represents a location-scale shift. We present results at the 0.25 and 0.75 quantiles. We find that the bias of the estimator in the simulations is very low, ranging from 0.7 percent to 5.0 percent in absolute value. The performance of the QRIE estimator continues to be satisfactory, offering in general the smallest bias and best MSE in the class of fixed effects estimators presented in the

table. Lastly, we expand the previous designs in Table 4.3 by replacing equation (4.1) by:

$$(4.6) \qquad y_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 x_t + c_1(\lambda_{1i} f_{1t} + \lambda_{2i} f_{2t}) + c_2 \lambda_{3i} + (1 + h d_{it}) u_{it},$$

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

> **Design 4:** The dependent variable is generated from the classical model with individual effects. We assume that $c_1 = 0$ and $c_2 = 1$.
>
> **Design 5:** We introduce a location-scale shift model with individual and interactive effects. We assume that $c_1 = c_2 = 1$.
>
> **Design 6:** The dependent variable is generated from a location-scale shift model with an endogenous variable, d. It is assumed that $\Omega_{21} = \Omega_{12} = 0.5$. The model also includes individual and interactive effects, because $c_1 = c_2 = 1$.

As before, Table 4.3 shows results obtained from the panel data quantile regression estimators introduced in Table 4.1. We also show results obtained from an extension of the QRIE estimator. The estimator, which is labelled as QRIIE, includes individual effects and interactive effects.

Table 4.3 presents the bias and root mean square error in panels of sample size $N = 100$ and $T = \{4, 6, 12, 20\}$. As expected, the performance of the QRFE estimator is satisfactory in the model with individual effects. We note, however, that the bias of the QRFE estimator tends to disappear as T increases. On the other hand, the QRFE has a poor MSE performance in all the variants of these models with interactive effects. When the model includes endogenous covariates, individual and interactive effects, the QRIIE estimator offers the best small sample performance in terms of bias and MSE.

Overall, the finite sample performances of the methods for models with interactive effects is very good in all the variants of the models considered in Tables 4.1-4.3. 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.

5. Educational outcomes and heterogeneous class size effects

In the last half a century, understanding what drives students' academic performance has been a major focus in the economics of education. The analysis of class size reduction on educational attainment continues to be one of the controversial topics in the social sciences ever since the

Coleman Report (1966). 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).[3]

We apply our quantile regression interactive effects method to a structural equation model of students' educational achievement where teachers' and students' unobserved heterogeneity are allowed to interact. We will then compare the policy implications of changes in the number of students in a class and changes in the class composition on educational achievement, obtained from a series of instrumental variable and panel data methods. We find that our method suggests different policies relative to other existing estimation approaches. The instrumental variable estimator suggests that a reduction in class size has an insignificant effect at the tails of the conditional distribution. We find, however, that while for the worst students, smaller classes improve performance, for the strongest students, smaller classes reduce performance. This is consistent with current research in educational psychology which suggests that weaker students benefit from the increased attention they receive in smaller classes, while strong students flourish more in competitive environments and benefit from the numerous interactions that larger classes provide.

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

5.1. Data

We employ data from the administrative records of the Economics, Management, and Finance programs at Bocconi University. The university is an established higher education institution located in Milan, Italy. The administration extracted a comprehensive data set of college students including information on students' characteristics and outcomes. The data set includes information on course grades, background demographic and socioeconomic characteristics such us gender, family income, and pre-enrollment test scores. Additionally, the data set includes information on

---

[3]A number of recent studies have focused on class size and peer effects (e.g., Krueger 1999, Angrist and Lavy 1999, Hoxby 2000, Hanushek et al. 2003). The literature to date offers a large number of studies on the effect of class size on educational achievement of the average student, but few studies investigate its distributional effect. One of the few exceptions are Levin (2001), Ma and Koenker (2006), and more recently, Bandiera, Larcinese, and Rasul (2010).

Variable	Mean	Std. Dev.	Minimum	Maximum
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

TABLE 5.1. *Descriptive statistics.*

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 5.1 reports briefly the summary statistics of the variables used in this study. The average grade is 25.602, which can be associated to a $B+$ in the US grading system (De Giorgi, Pellizzari and Woolston, 2009). The average class includes 134 students, ranging from 90 to 158 students. The average number of assigned students to these classes is 133, and ranges from 89 to 160 students. The table indicates a small degree of self-selection into classes post-randomization which explains the difference between actual class enrollment and allocation.

5.2. 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:

$$(5.1) \qquad y_{ict} \;=\; \boldsymbol{d}'_{ct}\boldsymbol{\alpha} + \boldsymbol{x}'_{ict}\boldsymbol{\beta} + \boldsymbol{z}'_t\boldsymbol{\gamma} + \boldsymbol{f}'_{ct}\boldsymbol{\lambda}_i + u_{ict}$$

$$(5.2) \qquad \boldsymbol{d}_{ct} \;=\; \boldsymbol{w}'_{ct}\boldsymbol{\pi}_1 + \boldsymbol{x}'_i\boldsymbol{\pi}_2 + \boldsymbol{f}'_{ct}\boldsymbol{\lambda}_i + \boldsymbol{v}_{ict},$$

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, z. The factors are approximated by cross-sectional averages including the average grade for class c at year t and the average size of class c at year t.

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

It is natural to control for individual and class-cohort latent heterogeneity by imposing a linear additive structure $\lambda_i + f_{ct}$ as in Hanushek et al. (2003). The individual effect 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 $f'\lambda$ could represent the interaction between students' 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 a student's academic performance. At the same time, these shocks could also affect the number of students in a class, because they may affect a student's decision to stay in the assigned class. We model these potential interactions by letting the error term u to be correlated with v.
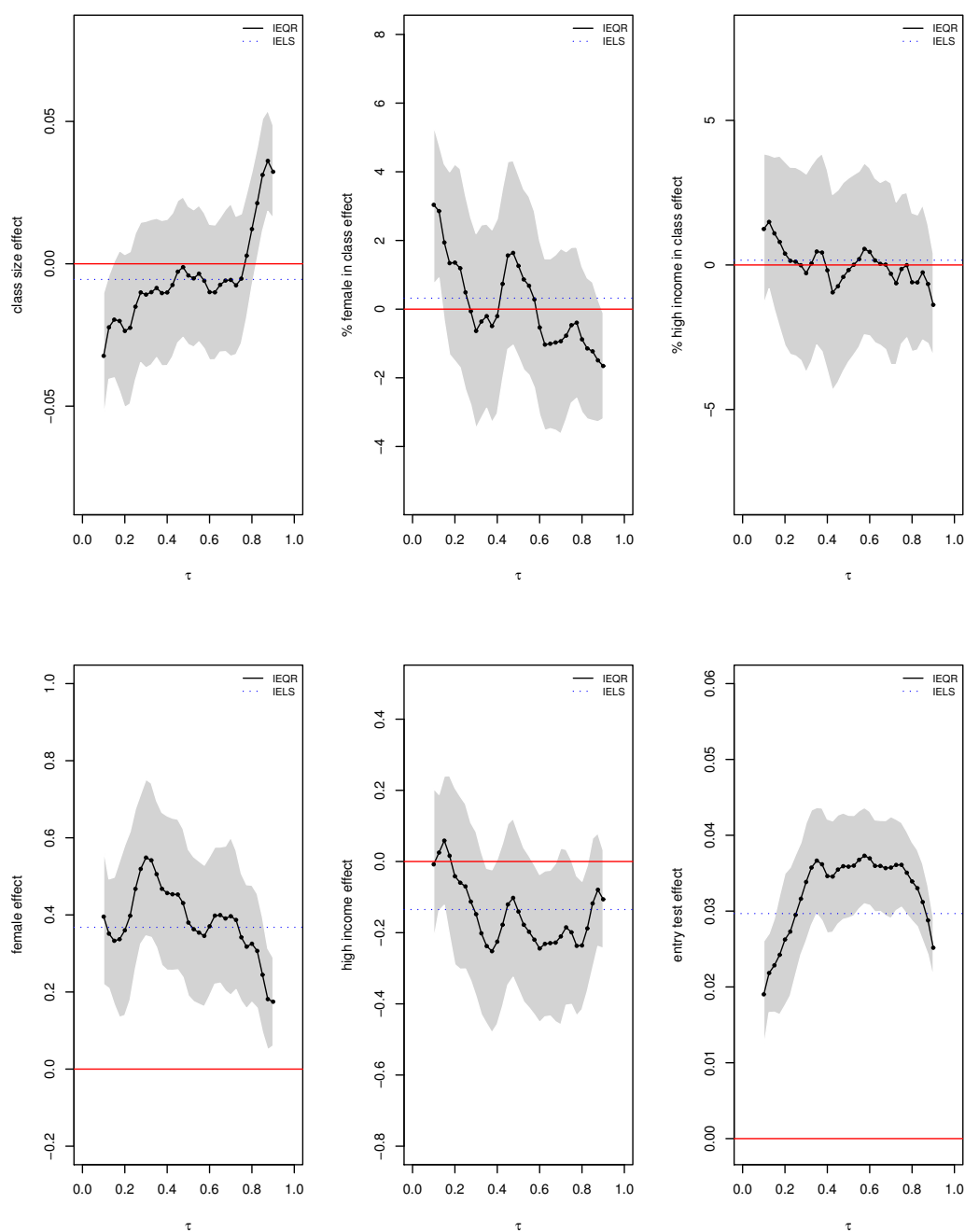
FIGURE 5.1. *Quantile regression covariate effects on educational attainment. The continuous dotted line shows quantile regression with interactive effects (QRIE) estimates and the dashed horizontal line shows the least squares version of the quantile estimator (LSIE).*

5.3. **An Empirical Analysis**

We apply the method proposed in equations (2.8)-(2.10) to the structural model (5.1)-(5.2). Figure 5.1 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. 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.[4]

Let us now consider the empirical evidence presented in Table 5.2. This table presents the estimated effects of the variable that may be of interest to policy makers: class size. 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

---

[4]Although the students were randomly assigned into classes, teachers were not. This issue is analyzed in De Giorgi, Pellizari and Woolston (2009), who find evidence that teachers' allocation is not a concern in this data. We investigated this potential issue by restricting the original sample to include classes of relatively similar size. We found that the effect of class size on academic performance suggests similar conclusions than before.

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	-0.034	-0.014	-0.005	-0.006	0.031	-0.005
	(0.014)	(0.016)	(0.015)	(0.014)	(0.010)	(0.010)

TABLE 5.2. *Panel data and IV estimates of the causal effect of class size on educational achievement. Standard errors are presented in parentheses.*

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

5.4. **Hypothesis Tests in a Model with Interactive Effects**

In this Section, we perform a Hausman-type test for exogeneity in a model with interactive effects and we briefly test if class size, the percent of females in class, and the percent of high income students in class are significantly different across quantiles. We test two hypotheses. First, we check whether the effects of interest are significantly different than zero at several quantiles of

Effects of Interest	Quantiles					0.1 and	
	0.1	0.25	0.5	0.75	0.9	0.50	0.90
Specification I: Equations (5.1)-(5.2)							
Class size effect	0.018	0.390	0.730	0.659	0.002	0.163	0.000
% Female in class	0.008	0.889	0.208	0.578	0.011	0.273	0.000
% High income in class	0.445	0.944	0.905	0.978	0.146	0.543	0.150
Interactive effects	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Specification II: Model with exogenous covariates and individual effects							
Class size effect	0.227	0.224	0.109	0.578	0.573	0.867	0.548
% Female in class	0.568	0.319	0.138	0.345	0.361	0.728	0.886
% High income in class	0.919	0.623	0.344	0.166	0.277	0.584	0.511
Interactive effects	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Specification III: Equation (5.1) with exogenous covariates							
Class size effect	0.000	0.000	0.000	0.000	0.000	-	-
% Female in class	0.003	0.784	0.446	0.766	0.001	-	-
% High income in class	0.018	0.768	0.191	0.907	0.182	-	-
Interactive effects	0.000	0.000	0.000	0.000	0.000	-	-

TABLE 5.3. *Specification tests in a quantile regression model with interactive effects. The table presents p-values.*

the conditional distribution of educational attainment. We also test if the interactive effect term is statistically significantly different than zero. Second, we test if the effects at the 0.1 quantile are significantly different than effects at other quantiles $\{0.50, 0.90\}$. These specification tests are performed on two models. First, we consider the model described in equations (5.1) and (5.2). We also introduce a variation to equation (5.1), assuming that the covariates are exogenous and that the model includes both fixed and interactive effects. This allows us to examine the importance of the interactive effect term in the classical fixed effects model for educational attainment. Lastly, we test for exogeneity in equation (5.1) accommodating the Wald test introduced in Section 3 into a Hausman-type test. We evaluate the hypothesis that $\boldsymbol{\alpha}(\tau) = \boldsymbol{a}(\tau)$, where $\hat{\boldsymbol{a}}(\tau)$ is a quantile regression estimator obtained from a model that controls for interactive effects but ignores the possible correlation between u_{it} and $\boldsymbol{v}_{it}$ in equations (5.1) and (5.2).

The results of several tests are presented in Table 5.3. While the upper block shows p-values of tests performed on our basic specification, the lower block presents p-values of tests performed on the variation of the model. We find that size effect are statistically different at the lower and upper tails, suggesting that changes in class size have a significant effect among low and high performers. We also see that changes in the gender composition of the class impact significantly different low- and high- performers. Moreover, it is interesting to see that the Wald tests reject

the null hypothesis of no interactive effects in the model with fixed effects (p-values are 0.000). We interpret this evidence as suggesting that the fixed effects specification is rejected in favor of the interactive effects specification.

The lower block of Table 5.3 presents results for a series of Hausman-type exogeneity tests. It is interesting to see that the tests reject the null hypothesis of exogeneity of class size in a model conditional on interactive effects. In contrast, the tests suggest that class composition represented by percentage of high-income students might be exogenous in our specification.

5.5. **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 $y(\tau_j) = q_{-c}(\tau_j) + \alpha_c(\tau_j)d_c$, for $\tau_j = \{\tau_L, \tau_H\}$. The function q_{-c} denotes the conditional quantile model corresponding to equation (5.1) 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 is split for simplicity in two classes, c and c'. It is straightforward to show that the solution has the following form:

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

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

Using Figure 5.2, 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 5.1. We observe that most of the point estimates are contained in the square in the left corner, suggesting that a small percentage
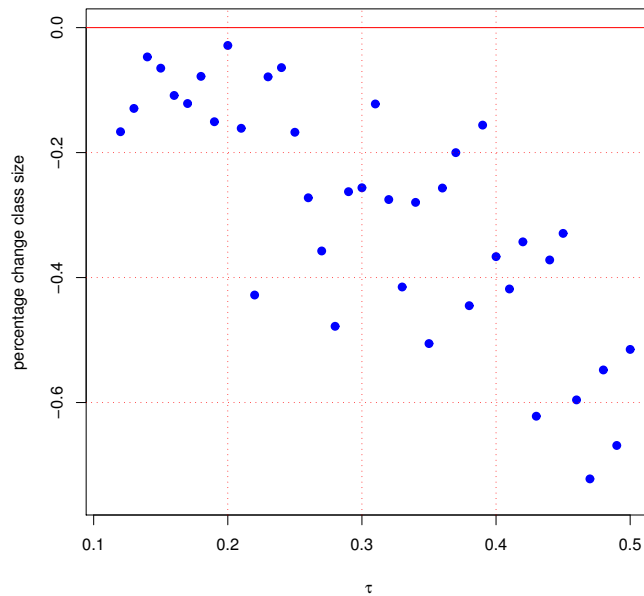
FIGURE 5.2. *Class size changes to prevent failure among the worst students.*

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.

6. Conclusion

This paper proposes a quantile regression estimator for a model with interactive effects potentially correlated with the independent variables and endogenous treatment effects. We provide conditions

under which the slope parameter estimator is consistent and asymptotically Gaussian. Monte Carlo studies are carried out to investigate the finite sample performance of the proposed method in comparison with other candidate methods. The evidence shows that the finite sample performance of the proposed method is excellent under different Monte Carlo designs. We also apply the new method to an investigation of the effect of class size on educational performance.

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

Appendix A. Proofs

Proof of Proposition 1. (Consistency) Under the regularity conditions, identification and consistency results immediately follows from the proof of Theorem 3 in Chernozhukov and Hansen (2006) and Corollary 3.2.3 in van der Vaart and Wellner (1996). By proposition 2 in Chernozhukov and Hansen (2008), we have that $\sup_{\alpha \in \mathcal{A}} \|\hat{\boldsymbol{\vartheta}}(\alpha, \tau) - \boldsymbol{\vartheta}(\alpha, \tau)\| \to 0$ for $\boldsymbol{\vartheta} = (\boldsymbol{\beta}', \boldsymbol{\delta}', \boldsymbol{\gamma}')'$. 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 $\boldsymbol{\alpha}(\tau)$. Then for any $\boldsymbol{\alpha}_n \to \boldsymbol{\alpha}(\tau)$, we have that $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \tau) \to \boldsymbol{\beta}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\beta}(\tau)$, $\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha}_n, \tau) \to \boldsymbol{\delta}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\delta}(\tau)$, and $\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_n, \tau) \to \boldsymbol{\gamma}(\boldsymbol{\alpha}(\tau), \tau) = \boldsymbol{\gamma}(\tau) = \mathbf{0}$. 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 $\rho_\tau(y_{it} - \boldsymbol{d}'_{it}\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{f}'_t\hat{\boldsymbol{\delta}}(\tau) - \boldsymbol{w}'_{it}\hat{\boldsymbol{\gamma}}(\tau))$ 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 $\xi_{it}(\tau) = \boldsymbol{d}'_{it}\boldsymbol{\alpha}(\tau) + \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau) + \boldsymbol{f}'_t\boldsymbol{\delta}(\tau) + \boldsymbol{w}'_{it}\boldsymbol{\gamma}(\tau)$, $\hat{\boldsymbol{\delta}}_\alpha(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\alpha}(\tau))$, $\hat{\boldsymbol{\delta}}_\beta(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\beta}(\tau))$, $\hat{\boldsymbol{\delta}}_\lambda(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\delta}}(\boldsymbol{\alpha}_n, \tau) - \boldsymbol{\delta}(\tau))$, and $\hat{\boldsymbol{\delta}}_\gamma(\boldsymbol{\alpha}_n, \tau) = \sqrt{TN}(\hat{\boldsymbol{\gamma}}(\boldsymbol{\alpha}_n, \tau) - \mathbf{0})$. Under assumption 4, the solution of (2.9) 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 $u_{it}(\tau) = y_{it} - \xi_{it}(\tau)$. Let,

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

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

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

Taking expectation and expanding v under condition 6, we obtain

$$\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})) =$$

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

$$\left. + \frac{1}{\sqrt{TN}} \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{f}_t \psi_\tau(u_{it}(\tau)) \right)$$

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

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 $\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma, \boldsymbol{\delta}_\lambda) - v(\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})) = v(\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})$. This last expression can be written as,

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

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,

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

$$\left. + \sum_{i=1}^N \sum_{t=1}^T \boldsymbol{f}_t \psi_\tau(u_{it}(\tau)) \right) + \frac{R_{it}}{\sqrt{TN}}$$

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

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,

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

Expanding as above we obtain

$$
\mathbb{E}(v(\boldsymbol{\delta}_\alpha, \boldsymbol{\delta}_\beta, \boldsymbol{\delta}_\gamma) - v(\mathbf{0},\mathbf{0},\mathbf{0})) =
$$

$$
= -\frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{h}_{it} g_{it}(\xi_{it}(\tau)) \left(\boldsymbol{d}'_{it} \frac{\boldsymbol{\delta}_\alpha}{\sqrt{NT}} + \boldsymbol{x}'_{it} \frac{\boldsymbol{\delta}_\beta}{\sqrt{TN}} + \boldsymbol{w}'_{it} \frac{\boldsymbol{\delta}_\gamma}{\sqrt{NT}} \right.
$$

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

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 $\boldsymbol{\delta}_\vartheta = (\boldsymbol{\delta}'_\beta, \boldsymbol{\delta}'_\gamma)'$, we write the last expression as,

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

$$
\frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} \boldsymbol{h}_{it} \psi_\tau \left(u_{it}(\tau) \right) - \frac{R_{it}}{\sqrt{NT}}
$$

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, $\hat{\boldsymbol{\delta}}_\gamma = \bar{\boldsymbol{J}}'_\gamma (\mathbb{J}_\psi - \boldsymbol{J}_\alpha \boldsymbol{\delta}_\alpha)$, and $\hat{\boldsymbol{\delta}}_\beta = \bar{\boldsymbol{J}}'_\beta (\mathbb{J}_\psi - \boldsymbol{J}_\alpha \boldsymbol{\delta}_\alpha)$.

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,

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

$\square$

References

ABREVAYA, J., AND C. DAHL (2008): "The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Regression Estimation on Panel Data," *Journal of Business and Economics Statistics*, 26(4), 379–397.

ANDO, T., AND R. S. TSAY (2010): "Quantile Regression Models with Factor-Augmented Predictors and Information Criterion," *Econometrics Journal, forthcoming.*

ANGRIST, J., AND V. LAVY (1999): "Using Maimonides' Rule to Estimate The Effects of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

BAI, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.

BALTAGI, B. (2005): *Econometric Analysis of Panel Data.* Wiley, New York, 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.*

CANAY, I. (2010): "A Note on Quantile Regression for Panel Data Models," mimeo, Northwestern University.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND W. NEWEY (2009): "Quantile and Average Effects in Nonseparable Panel Models,," cemmap Working Papers, CWP29/09.

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–262.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491–525.

——— (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379–398.

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

DE GIORGI, G., M. PELLIZZARI, AND W. WOOLSTON (2009): "Class-size and Class Heterogeneity," mimeo, Stanford University.

DINARDO, J., AND D. S. LEE (2010): "Program Evaluation and Research Designs," NBER WP 16016.

DOKSUM, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *Annals of Statistics*, 2, 267–277.

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.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–1271.

HECKMAN, J. J., AND E. VYTLACIL (2001): "Policy-Relevant Treatment Effects," *American Economic Review*, 91(2), 107–111.

HOXBY, C. (2000): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics*, 115, 1239–1285.

HSIAO, C. (2003): *Analysis of Panel Data*. Cambridge University Press, New York, 2nd edn.

KATO, K., AND A. F. GALVAO (2010): "Smoothed Quantile Regression for Panel Data," mimeo, Hiroshima University.

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

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

———— (2010): "Quantreg," R package version 4.48, www.r-project.org.

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

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

KOENKER, R., AND Z. XIAO (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114, 497–532.

LAMARCHE, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.

LEHMANN, E. (1974): *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day Inc.

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.

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

PESARAN, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

PONOMAREVA, M. (2010): "Quantile Regression for Panel Data Models with Fixed Effects and Small T: Identification and Estimation," mimeo, University of Western Ontario.

POWELL, D. (2009): "Unconditional Quantile Regression for Panel Data with Exogenous or Endogenous Regressors," Rand Working Paper WR-710.

ROSEN, A. (2009): "Set Identification via Quantile Restrictions in Short Panels," CEMMAP, WP 26/09.

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

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer, New York.

WEI, Y., AND X. HE (2006): "Conditional Growth Charts," *Annals of Statistics*, 34(5), 2069–2097.