

Instrumental Variables Quantile Regression for Panel Data with Measurement Errors*

Antonio F. Galvao, Jr.[†]
University of Illinois

Gabriel V. Montes-Rojas[‡]
City University London

January, 2009

Abstract

This paper develops an instrumental variables estimator for quantile regression in panel data with fixed effects. Asymptotic properties of the instrumental variables estimator are studied for large N and T when $N^a/T \rightarrow 0$, for some $a > 0$. Wald and Kolmogorov-Smirnov type tests for general linear restrictions are developed. Monte Carlo simulations are conducted to evaluate the finite sample properties of the estimator in terms of bias and root mean squared error. The estimator is applied to the problem of measurement errors in variables. We derive an approximation to the bias in the quantile regression fixed effects estimator and show its connection to similar effects in standard least squares models. Finally, the methods are applied to a model of firm investment. The results show interesting heterogeneity in the Tobin's q and cash flow sensitivities of investment. In both cases, the sensitivities are monotonically increasing along the quantiles.

Key Words: Quantile Regression; Panel Data; Measurement Errors; Instrumental Variables

JEL Classification: C14, C23

*The authors would like to express their appreciation to Roger Koenker, Heitor Almeida, Murillo Campello, and participants in the seminars at Cass Business School, University of Alicante, and 18th Midwest Econometrics Group Meeting for helpful comments and discussions. All the remaining errors are ours.

[†]Department of Economics, University of Illinois at Urbana-Champaign, 419 David Kinley Hall, 1407 W. Gregory Dr., Urbana, IL 61801, USA. Email: galvao@illinois.edu

[‡]Department of Economics, City University London, D308 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK. Email: Gabriel.Montes-Rojas.1@city.ac.uk

1 Introduction

Semiparametric panel data models have attracted interest in both theory and applications since the Gaussian assumptions underlying classical least-squares (LS) methods are sometimes implausible. Koenker (2004) introduced a general approach to estimation of quantile regression (QR) models for longitudinal data. Individual specific fixed effects (FE) are treated as pure location shift parameters common to all conditional quantiles and may be subject to shrinkage toward a common value as in the Gaussian random effects paradigm. QR methods allow one to explore a range of conditional quantiles exposing a variety of forms of conditional heterogeneity under weaker distributional assumptions and also provide a framework for robust estimation and inference. Controlling for individual specific heterogeneity via FE while exploring heterogeneous covariate effects within the QR framework offers a more flexible approach to the analysis of panel data than that afforded by the classical Gaussian fixed and random effects estimators. Recent work by Lamarche (2006, 2008) and Geraci and Bottai (2007) have elaborated on these models using penalized QR estimators. Abrevaya and Dahl (2007) have introduced an alternative approach to estimating QR models for panel data employing the “correlated random effects” model of Chamberlain (1982).

Economic quantities are frequently measured with error, particularly if longitudinal information is collected through one-time retrospective surveys, which are notoriously susceptible to recall errors. If the regressors are indeed subject to measurement errors (ME), it is well known that the slope coefficient of the least squares (LS) regression estimator is inconsistent because the measurement error induces endogeneity in the model. In the one regressor case (or the multiple regressor case with uncorrelated regressors), under standard assumptions, the ordinary LS estimator is biased toward zero, a problem often denoted as attenuation. The most common remedy to reduce this bias caused by the endogeneity problem is to use either economic theory or intuition to find additional observable variables that can serve as instrumental variables (IV). Most of the literature on the estimation of fixed effects (FE) in panel data models with measurement errors is based on LS with instrumental variables and the generalized method of moments (GMM). See for instance Hsiao (2003, 1992), Wansbeek (2001), Biorn (2000, 1992), Hsiao and Taylor (1991), Wansbeek and Koning (1989) and Griliches and Hausman (1986).

The goal of this paper is to contribute to this literature by developing an IV estimator for

QR with FE (IVQRFE hereafter) to reduce the bias caused by the presence of endogeneity in the model. In particular, the IV estimator for QR introduced by Chernozhukov and Hansen (2006, 2008) will be adapted to the panel data setting of Koenker (2004). We derive consistency and asymptotic distribution of this estimator and suggest a Wald type test for general linear hypotheses and a Kolmogorov-Smirnov test for linear hypotheses over a range of quantiles $\tau \in \mathcal{T}$, and derive the respective limiting distributions. Although the proposed estimator is general for endogeneity in QR with FE, we focus on the application ME because this is an important class of endogeneity problem in panel data from the applied research standpoint, and also, under certain conditions, the instruments are available within the model. In particular, if the latent regressor is autocorrelated or non-stationary, lags of the regressors are valid instruments for the mismeasured variable, and estimators of the coefficient of the latent regressor will be consistent. Based on Griliches and Hausman (1986), and Biorn (2000) the IV approach employs lagged (or lagged differences of the) regressors as instruments. Thus, the estimator combines the usual concept for panel data with ME and the IV QR framework.

Recently, the topic of ME in variables has also attracted considerable attention in the QR literature. Chesher (2001) studies the impact of covariate ME on quantile functions using small variance approximation, and Schennach (2008) discusses identification and estimation issues for general quantile functions based on Fourier transforms and previous results for nonlinear models (see Schennach, 2004, 2008). As a by-product contribution of the paper, we derive an approximation to the bias in the QR FE estimator in the presence of a mismeasured variable. The effect of ME on the slope coefficient estimators can be seen as an application of Angrist, Chernozhukov, and Fernandez-Val (2006) omitted variables formula, which provides a useful generalization of the problem of ME bias discussed above to the QR framework. This representation provides an explicit formulation for the bias in the slope coefficients and complements the results in Chesher (2001).

As first noted by Neyman and Scott (1948), leaving the individual heterogeneity unrestricted in a nonlinear or dynamic model generally results in inconsistent estimators of the common parameters due to the incidental parameters problem; that is, noise in the estimation of the FE when the time dimension is short results in inconsistent estimates of the common parameters due to the nonlinearity of the problem. QR estimation of panel data models with FE suffers from similar problems to those seen in the LS case when T is modest. Reliance on the existing bias reduction strategies using LS differencing, either temporally,

or via the usual deviation from individual means (within) transformation, is unsatisfactory in the QR setting. Linear transformations that are completely innocuous in the context of conditional mean models are highly problematic in the conditional quantile models since they alter in a fundamental way what is being estimated. Expectations enjoy the convenient property that they commute with linear transformations; quantiles do not.¹ Fortunately, there is no need to transform the QR model to compute the FE estimator. Even when the number of FE is large, interior point optimization methods using modern sparse linear algebra make direct estimation of the QR model quite efficient. Koenker (2004) overcomes this problem by using a large N and T asymptotics with the restriction that $N^a/T \rightarrow 0$, for some $a > 0$.²

Monte Carlo simulations are used to assess the properties of the estimators in terms of bias, root mean squared error, and to compare its performance with LS methods. The simulations show that the QRFE estimator is severely biased in the presence of ME, while the proposed IV estimator is able to reduce the bias even in short panels. In addition, the Monte Carlo experiments suggest that IVQRFE performs better than LS-IV in terms of bias and root mean squared error for non-Gaussian heavy-tailed distributions.

Finally, we provide an example to illustrate the new methods. We apply the IVQRFE estimator to study the relationship between investment and Tobin's q and cash flow, a topic of considerable attention in corporate finance and applied econometrics in general. Tobin's q is the ratio of the market valuation of a firm and the replacement value of its assets. Economic theory predicts that firms with a high value of q are attractive investment opportunities, whereas a low value of q indicates the opposite. As argued in Erickson and Whited (2000), the poor performance of the Tobin's q theory is probably due to the measurement error of q values. Related to this is the debate on whether cash flow has an effect on investment (Alti, 2003; Almeida, Campello, and Weisbach, 2004). We argue that QR methods capture important heterogeneity across firms in terms of their q and cash flow sensitivities of investment. In the upper conditional quantiles of investment, investment may be driven by

¹This intrinsic difficulty has been recognized by Abrevaya and Dahl (2007), among others, and is clarified by Koenker and Hallock (2000, p.19): "Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects."

²More recently, a number of additional approaches have been proposed to reduce the bias in dynamic and nonlinear panels. These methods use asymptotic approximations derived as both the number of individuals, N , and the number time series, T , go to infinity jointly; see, for example, Arellano and Hahn (2005) for a survey, and Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Hahn and Newey (2004), Bester and Hansen (2007), for specific approaches.

insiders knowledge of business opportunities or a particular capital structure that requires more investment, and therefore, we expect that investment would be more responsive to changes in q and cash flow, as the firms would use all additional resources to finance its projects. Moreover, high investment ratios reduce the marginal productivity of additional investment, and therefore, it would be difficult to get additional resources. Then, we expect that q and cash flow sensitivity will be higher in the upper quantiles than in the lower quantiles of investment. The results show interesting heterogeneity in the Tobin's q and cash flow sensitivities of investment. In both cases, the sensitivities are monotonically increasing along the quantiles. We show that q values are subject to measurement error, and we use past values of q in differences as instrumental variables to reduce the bias.

The rest of the paper is organized as follows. Section 2 presents the instrumental variables quantile regression panel data with FE estimator and derives its asymptotic properties and related inference. Section 4 has Monte Carlo experiments. In Section 5 illustrates the new approach in the Tobin's q theory of investment. Finally, Section 6 concludes the paper.

2 The Model and Assumptions

2.1 Measurement error bias in quantile regression

Consider the following representation of a panel data model with individual FE and ME

$$y_{it} = d'_{it}\eta + x'^*_{it}\beta + u_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (1)$$

where y_{it} is the response variable, d_{it} denotes a dummy variable that identifies the N distinct individuals in the sample, η denotes the N -vector of individual FE, and x^*_{it} is a $\dim(x)$ -vector of the mismeasured regressors. Suppose that we do not observe x^*_{it} , but rather x_{it} , which is a noisy measure of x^*_{it} subject to an additive ME³ ϵ_{it} ,

$$x_{it} = x^*_{it} + \epsilon_{it}. \quad (2)$$

It is assumed that ϵ_{it} is independent and identically distributed (*iid*) with $Var[\epsilon_{it}] = \sigma_\epsilon^2 < \infty$. Using equation (2) we can express (1) in terms of the observed y 's and x 's as

$$y_{it} = d'_{it}\eta + x'_{it}\beta + u_{it} - \epsilon'_{it}\beta \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (3)$$

³We do not consider Schennach (2008) case of a nonseparable error structure, rather we restrict ourselves to the traditional framework of additivity.

It follows that the observed regressor x_{it} in (3) will be correlated with the composite error, $u_{it} - \epsilon_{it}\beta$, inducing endogeneity in the model. This problem is of practical significance since the resulting bias may be large.

In the following paragraphs, we derive an approximation to the bias in the QR estimator in the presence of ME using Angrist, Chernozhukov, and Fernandez-Val (2006) (denoted ACFV hereafter) approximation to the omitted variable bias in QR. This approach shows that, as in LS, ME bias in QR can be derived analytically as an endogeneity bias, and provides a simpler representation than that in Chesher (2001). Since this is the first attempt in the QR literature to do so in this way, we do not tackle simultaneously the incidental parameter problem, but rather we consider a standard QR structure.⁴ The reader may skip these and go to section 2.2 for the asymptotic results of the paper.

The standard result for the LS estimator with ME can be seen as an omitted variables problem, where $-\epsilon_{it}$ is the omitted variable. In the rest of this section we omit the indexes i and t to simplify the notation. Define $v^* = [d', x^{*'}]'$, $\Lambda_\epsilon = [0', \epsilon']'$, $v = [d', x']' = v^* + \Lambda_\epsilon$ and $\varphi = [\eta', \beta']'$. In addition, let

$$\varphi^* = \underset{\varphi}{\operatorname{argmin}} E[y - v^{*\prime}\varphi]^2, \quad (4)$$

and

$$\varphi^\circ = \underset{\varphi}{\operatorname{argmin}} E[y - v'\varphi]^2, \quad (5)$$

where φ^* and φ° are the parameters that solve the population minimization problem. Under known conditions, of course, these are the probability limit of the corresponding estimators. However, we follow ACFV exposition as close as possible in this sub-section.

In this case,

$$\varphi^\circ = \varphi^* - (E[vv'])^{-1} E[v\epsilon'\beta^*] = \varphi^* - (E[v^*v^{*\prime} + \Lambda_\epsilon\Lambda_\epsilon'])^{-1} E[\Lambda_\epsilon\epsilon'\beta^*]. \quad (6)$$

This shows the standard result that the bias depends on the noise to signal ratio. Therefore, after some algebra, the bias in the least squares FE estimator of β can be also written as

$$\beta^\circ = \beta^* \left[1 - \frac{\sigma_\epsilon^2}{\operatorname{Var}(x_{it}^* - \bar{x}_i^*)} \right],$$

⁴Alternatively the results presented here can be seen as the special case where N is fixed and T goes to infinity.

where \bar{x}_i^* is the defined as $\frac{1}{T} \sum_{t=1}^T x_{it}^*$.

Consider now the τ th conditional quantile function of the response y ,

$$Q_y(\tau|d, x^*) = d'\eta^*(\tau) + x^{*\prime}\beta^*(\tau). \quad (7)$$

In this formulation η , and β are allowed to depend upon the quantile, τ , of interest.⁵ Using equation (2) one can rewrite (7) as

$$Q_y(\tau|v^*) = d'\eta^*(\tau) + (x - \epsilon)'\beta^*(\tau) = d'\eta^*(\tau) + x'\beta^*(\tau) - \epsilon'\beta^*(\tau) = Q_y(\tau|v, \epsilon). \quad (8)$$

As in the standard LS case with measurement error, the QR counterpart can be seen as an omitted variable problem, where ϵ is the omitted variable. We derive the approximate bias using the ACFV omitted variable bias formula.

Applying the standard estimation procedure in QR to equation (7) would solve

$$\varphi^*(\tau) = \underset{\varphi}{\operatorname{argmin}} E[\rho_\tau(y - v'\varphi)], \quad (9)$$

where $\rho_\tau(u) := u(\tau - I(u < 0))$ as in Koenker and Bassett (1978), which is analogous to equation (4). However, the QR FE estimator, as in the LS case, is biased in the presence of the mismeasured variables. In this case, in the problem of solving (8) omitting $-\epsilon$, the standard QR solves

$$\varphi^\circ(\tau) = \underset{\varphi}{\operatorname{argmin}} E[\rho_\tau(y - v'\varphi)]. \quad (10)$$

Here $\varphi^*(\tau)$ and $\varphi^\circ(\tau)$ are the parameters that solve the population minimization problem, defined in an analog way to ACFV paper.

The following Lemma shows that the measurement error bias in QR can be approximated to an expression similar to that in (6).

Lemma 1 *Assume that: (i) the conditional density function $f_y(y|v, \epsilon)$ exists and is bounded a.s.; (ii) $E[y]$, $E[Q_y(\tau|v, \epsilon)^2]$, and $E\|v', \epsilon'\|^2$ are finite; (iii) $\varphi^*(\tau)$ and $\varphi^\circ(\tau)$ uniquely solves equations (9) and (10) respectively; (iv) ϵ is independent of (d, x^*, u) . Then,*

$$\varphi^\circ(\tau) = \varphi^*(\tau) - (E[\omega_\tau(v, \epsilon) \cdot (v v')])^{-1} E[\omega_\tau(v, \epsilon) \cdot v \epsilon' \beta^*(\tau)]. \quad (11)$$

where $\omega_\tau(v, \epsilon) := \int_0^1 f_{u(\tau)}(u \cdot \Delta_\tau(v, \epsilon; \varphi^\circ(\tau)) | v, \epsilon) du / 2$ is a weighting function, and $\Delta_\tau(v, \epsilon; \varphi^\circ(\tau)) = v' \cdot (\varphi^\circ(\tau) - \varphi^*(\tau))' + \epsilon' \beta^*(\tau)$ is the QR specification error.

⁵This approach is different from that in Koenker (2004) where the individual effects are constant across quantiles. A straightforward modification of our estimator returns the constant individual effects case. We discuss briefly the details in Appendix 0.

Proof. The proof follows ACFV results for partial QR and omitted variables bias (p.545–548). Since the conditional quantile function is linear, $Q_y(\tau|v, \epsilon) = v'\varphi^*(\tau) - \epsilon'\beta^*(\tau)$, where $\varphi^*(\tau)$ is defined as in (9). Then, the conditional QR model in equation (7) can be seen as one with $[v', \epsilon']'$ as covariates. Moreover, the conditional QR with measurement error $Q_y(\tau|v) = v\varphi^\circ(\tau)$, obtaining the coefficient $\varphi^\circ(\tau)$ from equation (10), can be seen as a model with omitted variable ϵ .

Recall that $\omega_\tau(v, \epsilon) := \int_0^1 f_{u(\tau)}(u \cdot \Delta_\tau(v, \epsilon; \varphi^\circ(\tau))|v, \epsilon) du/2$ where $f_{u(\tau)}(\cdot|\cdot)$ is the conditional density function of $u(\tau) := y - Q_y(\tau|v, \epsilon)$, and $\Delta_\tau(v, \epsilon; \varphi) := v'\varphi - Q_y(\tau|v, \epsilon)$ is the bias in the estimated quantile function for a given φ . Then,

$$\Delta_\tau(v, \epsilon; \varphi^\circ(\tau)) = d' \cdot (\eta^\circ(\tau) - \eta^*(\tau)) + x^{*'} \cdot (\beta^\circ(\tau) - \beta^*(\tau)) + \epsilon'\beta^\circ(\tau).$$

Under the stated assumptions, by Theorem 2 in ACFV, $\varphi^\circ(\tau)$ uniquely solves the equation

$$\varphi^\circ(\tau) := \operatorname{argmin}_\varphi E[\omega_\tau(v, \epsilon)\Delta_\tau^2(v, \epsilon; \varphi)].$$

Solving for $\varphi^\circ(\tau)$ we have

$$\varphi^\circ(\tau) = \varphi^*(\tau) - (E[\omega_\tau(v, \epsilon) \cdot (vv')])^{-1} E[\omega_\tau(v, \epsilon) \cdot v\epsilon'\beta^*(\tau)].$$

■

Note that the weighting function $\omega(\cdot)$ depends on both v and ϵ and it is a distinctive feature of QR when compared with LS case. However, it can be shown that the leading term in the QR bias has the same form as that in LS. In order to show this, assume that $f_y(y|v, \epsilon)$ has a first derivative in y that is bounded in absolute value by $\bar{f}'(v, \epsilon)$ and consider a Taylor expansion of the weights as in ACFV, p.546,

$$\omega_\tau(v, \epsilon) = 1/2 \cdot f_y(Q_\tau(y|v, \epsilon)|v, \epsilon) + \varsigma(v, \epsilon),$$

where

$$|\varsigma(v, \epsilon)| \leq 1/4 \cdot |\Delta_\tau(v, \epsilon; \varphi^\circ(\tau))| \cdot \bar{f}'(v, \epsilon).$$

Note that by independence of y and ϵ , $f_y(Q_\tau(y|v, \epsilon)|v, \epsilon) = f_y(Q_\tau(y|v^*)|v^*)$ (with first derivative bounded by $\bar{f}'(v^*)$). Then, when either $\Delta_\tau(v, \epsilon; \varphi^\circ(\tau))$ or $\bar{f}'(v^*)$ is small,

$$\omega_\tau(v, \epsilon) \approx 1/2 \cdot f_y(Q_\tau(y|v^*)|v^*).$$

Then, the ACFV weighted LS approximation to QR implies that

$$\varphi^\circ(\tau) \approx \varphi^*(\tau) - (E[f_y(Q_\tau(y|v^*)|v^*) (v^*v^{*'} + \Lambda_\epsilon \Lambda_\epsilon')])^{-1} E[f_y(Q_\tau(y|v^*)|v^*) \Lambda_\epsilon \epsilon' \beta^*(\tau)]. \quad (12)$$

It is important to note that a key factor in the coefficient bias approximation given by (12) is the conditional density function $f_y(y|v^*)$. In addition, we can compare equations (6) and (12), and as in the OLS case, the bias in the variable with measurement error parameter in the QR framework depends on the noise to signal ratio, but in the QR case, weighted by the conditional density function $f_y(y|v^*)$.

2.2 Estimation

We consider the panel data model with measurement errors exposed in eqs. (1) and (2), and we add a covariate z without measurement error to the model. We repeat the equations for convenience

$$y_{it} = d'_{it}\eta + x'_{it}\beta + z_{it}\alpha + u_{it} \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (13)$$

with

$$x_{it} = x^*_{it} + \epsilon_{it}. \quad (14)$$

As in the LS case, the bias in QR can be ameliorated through the use of IV, w , that affect the determination of x^* but are independent of both ϵ and u . If the latent regressor is auto-correlated, provided that some structure is imposed on the disturbances and measurement errors, instruments are available within the model. Examples of these are lagged values of the mismeasured variable (in levels or in differences) or other covariates. Following Biorn (2000) we use lags of the first differences of the mismeasured variable as IV. In this particular case, the bias can be solved without relying on additional exogenous variables that do not belong to the model. If this is not the case, the method is still valid but it relies on the availability of valid exogenous instruments.⁶ Other types of measurement errors can be accommodated by the use of appropriate IV, although we consider this simple ME structure to concentrate on the asymptotic properties of the IV estimator.

Consider now the τ th conditional quantile function of the response y ,

$$Q_y(\tau|d, x^*, z) = d'\eta(\tau) + x^{*'}\beta(\tau) + z'\alpha(\tau). \quad (15)$$

⁶In a related work, Schennach (2008) discusses identification and estimation for general quantile functions in the presence of ME in the regressors, establishing the availability of IV that enables the identification and the consistent estimation of nonparametric QR models. Our estimator has the advantage over Schennach (2008) estimator that it can be implemented using standard QR and IV techniques.

Following Chernozhukov and Hansen (2006, 2008), from the availability of IV w_{it} , we consider estimators defined as:

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \|\hat{\gamma}(\beta, \tau)\|_A, \quad (16)$$

where

$$(\hat{\eta}(\beta, \tau), \hat{\alpha}(\beta, \tau), \hat{\gamma}(\beta, \tau)) = \underset{\eta, \alpha, \gamma}{\operatorname{argmin}} \sum_{i=1}^N \sum_{t=1}^T \rho_{\tau}(y_{it} - d'_{it}\eta - x'_{it}\beta - z'_{it}\alpha - w'_{it}\gamma), \quad (17)$$

with $\|x\|_A = \sqrt{x'Ax}$ and A is a positive definite matrix.⁷ As suggested by Chernozhukov and Hansen (2008), with abuse of notation, in practice, a simple procedure is to let the instruments w_{it} either be w_{it} or the predicted value from a least squares projection of x_{it} on w_{it} and z_{it} . Our final parameter estimators are thus

$$\hat{\varphi}(\tau) := (\hat{\eta}(\hat{\beta}(\tau), \tau), \hat{\beta}(\tau), \hat{\alpha}(\hat{\beta}(\tau), \tau)). \quad (18)$$

The intuition underlying the estimator is that, since w is a valid instrument, it is independent of ϵ and u , and therefore, it should have a zero coefficient in (17). Thus, for given β , the QR of $(y_{it} - x'_{it}\beta)$ on the variables (d_{it}, z_{it}, w_{it}) should generate a zero coefficient for the variable w_{it} . Hence, by minimizing the coefficient of the variable w_{it} one can recover a consistent estimator of η , β and α . Values of lagged x and lags of the exogenous variable z affect the determination of x^* but are independent of ϵ , so they can be used as instruments.⁸ The instrumental variables quantile regression (IVQR) estimator method may be viewed as an appropriate QR analog of the two stage least squares (2SLS). The 2SLS estimates can be obtained by using the same two steps procedure as described above for the IVQR. In Appendix 1 we show the details of the derivation.

The instrumental variables quantile regression with fixed effects (IVQRFE) estimator can be implemented as follows:

1) For a given quantile of interest τ , define a grid of values $\{\beta_j, j = 1, \dots, J\}$, and run the ordinary τ -QR of $(y_{it} - x'_{it}\beta_j)$ on (d_{it}, z_{it}, w_{it}) to obtain coefficients $\hat{\eta}(\beta_j, \tau)$, $\hat{\alpha}(\beta_j, \tau)$ and

⁷As discussed in Chernozhukov and Hansen (2006), the exact form of A is irrelevant when the model is exactly identified, but it is desirable to set A equal to the asymptotic variance-covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$ otherwise.

⁸It is known that if the distribution of the latent regressor vector is not time invariant, then consistent IV estimators of the coefficient of the latent regressor vector exist (see e.g. Griliches and Hausman, 1986; Biorn, 2000). The consistency of these estimators is robust to potential correlation between the individual heterogeneity and the latent regressor. An interesting point is that serial correlation or non-stationarity of the latent regressor is favorable from the point of view of identification and estimability.

$\hat{\gamma}(\beta_j, \tau)$, that is, by solving equation (17) for each value β_j .

2) Find $\hat{\beta}(\tau)$ as the value among $\{\beta_j, j = 1, \dots, J\}$ that makes $\|\hat{\gamma}(\beta_j, \tau)\|_A$ the closest to zero. This gives us the parameters in (18). We shall show that this estimator is consistent and asymptotically normal under some regularity conditions.

Given the parameter estimates, for an individual i , the τ -th conditional quantile function of y_{it} , can be estimated by,

$$\hat{Q}_{y_{it}}(\tau|d_{it}, x_{it}, z_{it}) = d'_{it}\hat{\eta}(\tau) + x'_{it}\hat{\beta}(\tau) + z'_{it}\hat{\alpha}(\tau).$$

In addition, given a family of estimated conditional quantile functions for an individual i , the conditional density of y_{it} at several values of the conditioning covariate can be estimated by the difference quotients,

$$\hat{f}_{y_{it}}(\tau|d_{it}, x_{it}, z_{it}) = \frac{\tau_k - \tau_{k-1}}{\hat{Q}_{y_{it}}(\tau_k|d_{it}, x_{it}, z_{it}) - \hat{Q}_{y_{it}}(\tau_{k-1}|d_{it}, x_{it}, z_{it})},$$

for some appropriately chosen sequence of τ s.

Now we discuss the asymptotic properties of the IVQRFE estimator. The parameter η , whose dimension N is tending to infinity, raises some new issues for the asymptotic analysis of the proposed estimator. Following the recent literature on panel data, we derive consistency and asymptotic normality assuming that both N and T tend to infinity. We impose the following regularity conditions:

A1 The y_{it} are independent with conditional distribution functions F_{it} , and differentiable conditional densities, $0 < f_{it} < \infty$, with bounded derivatives f'_{it} for $i=1, \dots, N$ and $t=1, \dots, T$;

A2 Let $D = I_N \otimes \iota_T$, and ι_T a T -vector of ones, $X = (x_{it})$ be a $NT \times \dim(\beta)$ matrix, $Z = (z_{it})$ be a $NT \times \dim(\alpha)$ matrix, and $W = (w_{it})$ be a $NT \times \dim(\gamma)$ matrix. For

$$\begin{aligned} \Pi(\eta, \beta, \alpha, \tau) &:= E[(\tau - 1(D'\eta + X'\beta + Z'\alpha))\check{X}] \\ \Pi(\eta, \beta, \alpha, \gamma, \tau) &:= E[(\tau - 1(D'\eta + X'\beta + Z'\alpha + W'\gamma))\check{X}] \\ \check{X} &:= [D', W', Z']', \end{aligned}$$

Jacobian matrices $\frac{\partial}{\partial(\eta, \beta, \alpha)}\Pi(\eta, \beta, \alpha, \tau)$ and $\frac{\partial}{\partial(\eta, \alpha, \gamma)}\Pi(\eta, \beta, \alpha, \gamma, \tau)$ are continuous and have full rank, uniformly over $\mathcal{E} \times \mathcal{B} \times \mathcal{A} \times \mathcal{G} \times \mathcal{T}$. The parameter space, $\mathcal{E} \times \mathcal{A} \times \mathcal{B}$, is a connected set. Moreover the image of $\mathcal{E} \times \mathcal{A} \times \mathcal{B}$ under the map $(\eta, \alpha, \beta) \rightarrow \Pi(\eta, \alpha, \beta, \tau)$ is simply connected;

A3 Denote $\Phi(\tau) = \text{diag}(f_{it}(\xi_{it}(\tau)))$, where $\xi_{it}(\tau) = d'_{it}\eta(\tau) + x'_{it}\beta(\tau) + z'_{it}\alpha(\tau) + w_{it}\gamma(\tau)$, $M_D = I - P_D$ and $P_D = D(D'\Phi(\tau)D)^{-1}D'\Phi(\tau)$. Let $\tilde{X} = [Z, W]'$. Then, the following matrix is invertible:

$$J_{\alpha\gamma} = E(\tilde{X}'M_D\Phi(\tau)M_D\tilde{X});$$

Now define $[\bar{J}'_{\alpha}, \bar{J}'_{\gamma}]'$ as a partition of $J_{\alpha\gamma}^{-1}$, $J_{\beta} = E(\tilde{X}'M_D\Phi(\tau)M_D X)$ and $H = \bar{J}'_{\gamma}A[\beta(\tau)]\bar{J}_{\gamma}$. Then, $J'_{\beta}HJ_{\beta}$ is also invertible;

A4 For all $\tau \in \mathcal{T} = [c, 1 - c]$ with $c \in (0, 1/2)$, $(\beta(\tau), \alpha(\tau)) \in \text{int } \mathcal{B} \times \mathcal{A}$, and $\mathcal{B} \times \mathcal{A}$ is compact and convex;

A5 $\max_{it} \|x_{it}\| = O(\sqrt{NT})$; $\max_{it} \|z_{it}\| = O(\sqrt{NT})$; $\max_{it} \|w_{it}\| = O(\sqrt{NT})$;

A6 $\frac{N^a}{T} \rightarrow 0$, for some $a > 0$.

Condition A1 is a standard assumption in the QR literature and imposes a restriction on the density function of y_{it} . Condition A2 is important for the parameter's identification. This is shown through the use of a version of Hadamard's theorem, as discussed in Chernozhukov and Hansen (2006). It requires that the instrument w impacts the conditional distribution of y at many relevant points. In addition, the condition that the image of the parameter space be simply connected requires that the image can be continuously shrunk to a point, and this condition can be interpreted as ruling out "holes" in the image of the set.⁹ Assumption A3 states invertibility conditions for matrices in order to guarantee asymptotic normality. A4 imposes compactness on the parameter space of $\beta(\tau)$. Such assumption is needed since the objective function is not convex in β . Assumption A5 imposes a bound on the variables. Finally, condition A6 is the same assumption as in Koenker (2004) and allows T to grow very slowly relative to N .

To further comment on the nature of correlation between X and W required by A2, note that by A1 we have that

$$\partial E[(\tau - 1(D'\eta + X'\beta + Z'\alpha))\tilde{X}]/\partial(\eta, \beta, \alpha) = E[(D', Z', W)'\Phi(\tau)(D, X, Z)]$$

Hence, the Jacobian in A2 takes a form of density-weighted covariance matrix for D , X and W , and A2 requires that this matrix has full rank. In addition, A2 imposes that global identifiability must hold; hence, the impact of W should be rich enough to guarantee that the equations are solved uniquely.

⁹We assume that the image of $\mathcal{E} \times \mathcal{A} \times \mathcal{B}$ under the map $(\eta, \alpha, \beta) \rightarrow \Pi(\eta, \alpha, \beta, \tau)$ is connected for ease of exposition. However, it is straightforward to show that the image of a connected set by an continuous function is a connected set.

We can now establish consistency and asymptotic normality. Proofs appear in Appendix 2. The following theorem states identification and consistency of $\hat{\varphi}(\tau)$.

Theorem 1 *Define $\psi_\tau(u) := (\tau - I(u < 0))$. Given assumptions A1-A6, $(\eta(\tau), \beta(\tau), \alpha(\tau))$ uniquely solves the equations $E[\psi_\tau(Y - D'\eta - X'\beta - Z'\alpha)\check{X}] = 0$ over $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$, and $\varphi(\tau) = (\eta(\tau), \beta(\tau), \alpha(\tau))$ is consistently estimable.*

It is important to notice that even if the dimension of the parameter space increases with the number of cross-section, we only need to impose that parameter space, $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$, is simple connected rather than compact. Therefore, by applying a Hadamards global univalence theorem for general metric spaces it is possible to show that there is a one-to-one correspondence between the parameter space and $\Pi(\mathcal{E}, \mathcal{B}, \mathcal{A}, \tau)$, the image of $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$ under $\Pi(\cdot, \cdot, \cdot, \tau)$. In addition, the identification follows from the global convexity of the quantile function and the instrumental variables exclusion restriction.

Under conditions A1-A6 we show that the asymptotic properties of the IVQRFE estimator as $N^a/T \rightarrow 0$ for some $a > 0$. Theorem 1 provides a lower bound for the rate at which T grows relative to N for ensuring consistency. The intuition behind this condition is that T must go to infinity fast enough to guarantee consistent estimates for the FE.

The intuition behind the proof of consistency relies on the uniform convergence of the objective function over the parameter space. The basic technique used to show this uniform convergence is similar to Wei and He (2006) where we divide the growing parameter space into small cubes. Then the total number of cubes grows at a polynomial rate so that the exponential bound obtained at each cube holds globally and the uniform convergence follows. Using this technique we establish stochastic equicontinuity, and consistency follows from application of an argmax theorem as in van der Vaart and Wellner (1996).¹⁰

In general, η is only a nuisance parameter with no particular interest. Define $\theta := (\beta, \alpha)$ as the vector of parameters of interest in this context. The limiting distribution of these parameters' IVQRFE estimators is given by Theorem 2.

Theorem 2 (*Asymptotic Normality*)

¹⁰It is important to note that we can use this technique of proof since the parameter space is growing at a known rate, N . There is a large literature showing the asymptotic properties of QR estimator for infinite dimension parameter space as He and Shao (2000) and Portnoy (1985) when the rate of the parameter increasing is not known. However, it is possible to achieve a better rate of N relative to T in the first case.

Under conditions A1-A6, for a given $\tau \in (0, 1)$, $\hat{\theta}(\tau)$ converges to a Gaussian distribution as

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} N(0, \Omega(\tau)), \quad \Omega(\tau) = (K', L')'S(K', L')$$

where $S = \tau(1 - \tau)E[VV']$, $V = \tilde{X}'M_D$, $K = (J'_\beta H J_\beta)^{-1}J_\beta H$, $H = \bar{J}'_\gamma A[\beta(\tau)]\bar{J}_\gamma$, $L = \bar{J}'_\alpha M$, $M = I - J_\beta K$, $J_\beta = E(\tilde{X}'M_D\Phi(\tau)M_D\tilde{X})$, $[\bar{J}'_\alpha, \bar{J}'_\gamma]'$ is a partition of $J_{\alpha\gamma}^{-1} = (E(\tilde{X}'M_D\Phi M_D\tilde{X}))^{-1}$, $\Phi(\tau) = \text{diag}(f_{it}(\xi_{it}(\tau)))$, and $\tilde{X} = [W, Z]'$.

The proof of asymptotic normality has some elements of the nonlinear panel data literature where we concentrate out the FE. Here we write an asymptotic representation for the FE, then we plug them into the representation for all the coefficients. Therefore, there will be a reminder term coming from this two step procedure. It happens that the large N and T asymptotics with the restriction that $N^a/T \rightarrow 0$ is a sufficient condition to ensure that the reminder term is negligible and the estimator is asymptotically normal centered at zero.

Remark 1. For a finite collection of quantile indexes we have

$$\sqrt{NT}(\hat{\theta}(\tau_j) - \theta(\tau_j))_{j \in \mathcal{J}} \xrightarrow{d} N(0, \Omega(\tau_k, \tau_l))_{k, l \in \mathcal{J}}$$

where $\Omega(\tau_k, \tau_l) = (K(\tau_k)', L(\tau_k)')'S(\tau_k, \tau_l)(K(\tau_l)', L(\tau_l)')$.

Remark 2. When $\dim(\gamma) = \dim(\beta)$, the choice of $A(\beta)$ does not affect asymptotic variance, and the joint asymptotic variance of $\beta(\tau)$ and $\alpha(\tau)$ will generally have the simple form $\Omega(\tau) = (K', L')'S(K', L')$, for S, K and L as defined above. As in Chernozhukov and Hansen (2008), when $\dim(\gamma) > \dim(\beta)$, the choice of the weighting matrix $A(\beta)$ generally matters, and it is important for efficiency. A natural choice for $A(\beta)$ is given by the inverse of the covariance matrix of $\hat{\gamma}(\beta(\tau), \tau)$. Noticing that $A(\beta)$ is equal to $(\bar{J}'_\gamma S \bar{J}_\gamma)^{-1}$ at $\beta(\tau)$, it follows that the asymptotic variance of $\sqrt{NT}(\hat{\beta}(\tau) - \beta(\tau))$ is given by

$$\Omega_\beta = (J'_\beta \bar{J}'_\gamma (\bar{J}_\gamma S \bar{J}_\gamma)^{-1} \bar{J}_\gamma J_\beta)^{-1}.$$

The components of the asymptotic variance matrix that need to be estimated include $J_{\alpha\gamma}, J_\beta$ and S . The matrix S can be estimated by its sample counterpart

$$\hat{S}(\tau, \tau') = (\min(\tau, \tau') - \tau\tau') \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T V_{it} V'_{it}. \quad (19)$$

Following Powell (1986), $J_{\alpha\gamma}$ and J_β can be estimated as stated in Theorem 2 above, such as $J_{\alpha\gamma} = E(\tilde{X}'M_D\Phi(\tau)M_D\tilde{X})$ and $J_\beta = E(\tilde{X}'M_D\Phi(\tau)M_D\tilde{X})$. The estimator of $J_{\alpha\gamma}$ is given

by the following form

$$\hat{J}_{\alpha\gamma} = \frac{1}{2NT h_n} \sum_{i=1}^N \sum_{t=1}^T I(|\hat{u}(\tau)| \leq h_n) \tilde{X} M_D M_D' \tilde{X}' \quad (20)$$

where $\hat{u}(\tau) := Y - D\hat{\eta}(\tau) - X\hat{\beta}(\tau) - Z\hat{\alpha}(\tau)$ and h_n is an appropriately chosen bandwidth, with $h_n \rightarrow 0$ and $NT h_n^2 \rightarrow \infty$. The estimator of J_β is analogous to $\hat{J}_{\alpha\gamma}$. By using the same procedure we can estimate the element $D\hat{\Phi}(\tau)D$ in P_D . The consistency of these asymptotic covariance matrix estimators are standard and will not be discussed further in this paper.

2.3 Inference

In this section, we turn our attention to inference in the IVQRFE model, and suggest a Wald type test for general linear hypotheses, and a Kolmogorov-Smirnov test for linear hypothesis over a range of quantiles $\tau \in \mathcal{T}$.

In the independent and identically distributed setup the conditional quantile functions of the response variable, given the covariates, are all parallel, implying that covariates effects shift the location of the response distribution but do not change the scale or shape. However, slopes estimates often vary across quantiles implying that it is important to test for equality of slopes across quantiles. Wald tests designed for this purpose were suggested by Koenker and Bassett (1982a), Koenker and Bassett (1982b), and Koenker and Machado (1999). It is possible to formulate a wide variety of tests using variants of the proposed Wald test, from simple tests on a single QR coefficient to joint tests involving many covariates and distinct quantiles at the same time. General hypotheses on the vector $\theta(\tau)$ can be accommodated by Wald type tests. The Wald process and associated limiting theory provide a natural foundation for the hypothesis $R\theta(\tau) = r$, when r is known.

Following the arguments of Portnoy (1984) and Gutenbrunner and Jureckova (1992), the QR process is tight and thus the limiting variate viewed as a function of τ is a Brownian Bridge over $\tau \in \mathcal{T}$.¹¹ Therefore, under the linear hypothesis $H_0 : R\theta(\tau) = r$, conditions A1-A6, and letting $\Sigma = (K', L')' EVV'(K', L')$, we have

$$\mathcal{V}_{NT} = \sqrt{NT} [R\Sigma(\tau)R']^{-1/2} (R\hat{\theta}(\tau) - r) \Rightarrow B_q(\tau), \quad (21)$$

¹¹In a related result, Wei and He (2006) establish tightness of the QR process in the longitudinal data context with increasing parameter dimension, see for instance Lemma 8.4.

where $B_q(\tau)$ represents a q -dimensional standard Brownian Bridge. For any fixed τ , $B_q(\tau)$ is $N(0, \tau(1 - \tau)I_q)$. The normalized Euclidean norm of $B_q(\tau)$

$$Q_q(\tau) = \|B_q(\tau)\|/\sqrt{\tau(1 - \tau)}$$

is generally referred to as a Kiefer process of order q . Thus, for given τ , the regression Wald process can be constructed as

$$\mathcal{W}_{NT}(\tau) = NT(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r) \quad (22)$$

where $\hat{\Omega}$ is a consistent estimator of Ω , and Ω is given by

$$\Omega(\tau) = (K'(\tau), L'(\tau))'S(\tau, \tau)(K'(\tau), L'(\tau)).$$

Under H_0 , the statistic \mathcal{W}_{NT} is asymptotically χ_q^2 with q -degrees of freedom, where q is the rank of the matrix R . The limiting distributions of the test is summarized in the following theorem:

Theorem 3 (*Wald Test Inference*). *Under $H_0 : R\theta(\tau) = r$, for fixed τ and conditions A1-A6,*

$$\mathcal{W}_{NT}(\tau) \stackrel{a}{\sim} \chi_q^2.$$

Proof. The proof of Theorem 3 is very simple and it follows from observing that for any fixed τ , by Theorem 2

$$\sqrt{NT}(\hat{\theta}(\tau) - \theta(\tau)) \Rightarrow N(0, \Omega(\tau))$$

under the null hypothesis,

$$\sqrt{NT}(R\hat{\theta}(\tau) - r) \Rightarrow N(0, R\Omega(\tau)R')$$

since $\hat{\Omega}(\tau)$ is a consistent estimator of $\Omega(\tau)$, by Slutsky theorem

$$\mathcal{W}_{NT}(\tau) = NT(R\hat{\theta}(\tau) - r)'[R\hat{\Omega}(\tau)R']^{-1}(R\hat{\theta}(\tau) - r) \stackrel{a}{\sim} \chi_q^2.$$

■

In order to implement the test it is necessary to estimate $\Omega(\tau)$ consistently. It is possible to obtain such estimator as suggested in Theorem 2 in the previous section, and the main components of $\hat{\Omega}(\tau)$ can be obtained as in equations (19) and (20).

More general hypotheses are also easily accommodated by the Wald approach. Let $v = (\theta(\tau_1)', \dots, \theta(\tau_m)')$ and define the null hypothesis as $H_0 : Rv = r$. The test statistic is the same Wald test as equation (22). However, Ω is now the matrix with kl th block

$$\Omega(\tau_k, \tau_l) = (K'(\tau_k), L'(\tau_k))' S(\tau_k, \tau_l) (K'(\tau_l), L'(\tau_l)),$$

and $S(\tau_k, \tau_l) = (\tau_k \wedge \tau_l - \tau_k \tau_l) E[VV']$, $\Phi(\tau_j) = \text{diag}(f_{it}(\xi_{it}(\tau_j)))$, and the other variables are as defined above. The statistic \mathcal{W}_{NT} is still asymptotically χ_q^2 under H_0 where q is the rank of the new matrix R . This formulation accommodates a wide variety of testing situations, from a simple test on single quantiles regression coefficients to joint tests involving several covariates and several distinct quantiles. Thus, for instance, we might test for the equality of several slope coefficients across several quantiles.

Another important class of tests in the QR literature involves the Kolmogorov-Smirnov (KS) type tests, where the interest is to examine the property of the estimator over a range of quantiles $\tau \in \mathcal{T}$, instead of focusing only on a selected quantile. Thus, if one has interest in testing $R\theta(\tau) = r$ over $\tau \in \mathcal{T}$, one may consider the KS type sup-Wald test. Following Koenker and Xiao (2006), we may construct a KS type test on the panel data regression quantile process in the following way:

$$KSW_{NT} = \sup_{\tau \in \mathcal{T}} \mathcal{W}_{NT}(\tau). \tag{23}$$

The limiting distribution of the Kolmogorov-Smirnov test is given in the following theorem:

Theorem 4 (*Kolmogorov-Smirnov Test*). *Under H_0 and conditions A1-A6,*

$$KSW_{NT} = \sup_{\tau \in \mathcal{T}} \mathcal{W}_{NT}(\tau) \Rightarrow \sup_{\tau \in \mathcal{T}} Q_q^2(\tau).$$

The proof of Theorem 4 follows directly from the continuous mapping theorem and equation (23). Critical values for $\sup Q_q^2(\tau)$ have been tabled by DeLong (1981) and, more extensively, by Andrews (1993) using simulation methods.

3 Monte Carlo Simulation

3.1 Monte Carlo design for the performance of IVQRFE

In this section, we describe the design of some simulation experiments that have been conducted to assess the finite sample performance of the IVQRFE estimator discussed in the previous sections.¹² Two simple versions of the basic model (1) are considered in the simulation experiment. In the first, reported in Tables 1 and 2, the scalar covariate, z_{it} , exerts a pure location shift effect. In the second, reported in Tables 3 and 4, z_{it} exerts both location and scale effects. In the former case the response y_{it} is generated by the model

$$y_{it} = \eta_i + x_{it}^* \beta + z_{it}' \alpha + u_{it},$$

while in the latter

$$y_{it} = \eta_i + x_{it}^* \beta + z_{it}' \alpha + (\gamma z_{it}) u_{it}.$$

In both cases we have an additive measurement error of the form

$$x_{it} = x_{it}^* + \epsilon_{it},$$

where x_{it}^* follows an ARMA(1,1) process

$$(1 - \phi L)x_{it}^* = \mu_i + \varepsilon_{it} + \theta \varepsilon_{it-1} \tag{24}$$

and ε_{it} follows the same distribution as u_{it} , that is, normal distribution and t_3 for Schemes 1 and 2 respectively. In all cases we set $x_{i,-50}^* = 0$ and generate x_{it}^* for $t = -49, -48, \dots, T$ according to

$$x_{it}^* = \mu_i + \phi x_{it-1}^* + \varepsilon_{it} + \theta \varepsilon_{it-1}. \tag{25}$$

This ensures that the results are not unduly influenced by the initial values of the x_{it}^* process. **Given the structure of the measurement error we use x_{it-1} as instrumental variable.**

We employ two different schemes to generate the disturbances $(u_{it}, \varepsilon_{it})$. Under Scheme 1, we generate them as $N(0, 1)$. Under Scheme 2 we generate them as t -distribution with 3 degrees of freedom. The exogenous regressor z_{it} is generated according to the same distribution as the innovations. In order to estimate the model we discarded the first 50 time series

¹²The experiment shown in this section is based in Hsiao, Pesaran, and Tahmiscioglu (2002).

observations, using the observations $t = 0$ through T for estimation. The FE, μ_i and η_i , are generated as

$$\mu_i = e_{1i} + T^{-1} \sum_{t=1}^T \varepsilon_{it}, \quad e_{2i} \sim N(0, \sigma_{e_2}^2).$$

$$\eta_i = e_{2i} + T^{-1} \sum_{t=1}^T z_{it}, \quad e_{2i} \sim N(0, \sigma_{e_2}^2).$$

The above method of generating μ_i and η_i ensures that the random effects estimator is inconsistent because of the correlation that exists among the individual effects and the explanatory variables.

In the simulations, we experiment with $T = 10$ and $N = 100$. We set the number of replications to 1000, and consider the following values for the remaining parameters:

$$(\beta, \alpha) = (1, 1);$$

$$\phi = 0.6, \quad \theta = 0.7, \quad \gamma = 1, \quad \sigma_u^2 = \sigma_{e_1}^2 = \sigma_{e_2}^2 = 1.$$

In the Monte Carlo study, we compare the LS and QR based estimators in terms of bias and root mean squared error (RMSE). More specifically, we study four estimators: the usual least squares within group estimator (FE); the least squares instrumental variables estimator (IVFE); the quantile regression fixed effects estimator proposed by Koenker (2004) (QRFE); and finally, the instrumental variables quantile regression fixed effects estimator proposed in this paper (IVQRFE). The QR estimators are studied only for the median case, although similar results are obtained for other quantiles.

3.2 Monte Carlo Results

Tables 1 and 2 present bias and RMSE simulations for estimates of the mismeasured coefficient, β , and the perfectly measured variable coefficient, α , for the location-shift model and the normal and t_3 distributions of the innovations respectively. The results show that the bias in the mismeasured variable coefficient of the FE and QRFE estimators are severe. However the coefficients for the IVFE and IVQRFE are approximately unbiased. Moreover, since we do not consider any correlation among the regressors, as expected, the coefficient of the perfectly measured variable is approximately unbiased in all cases.

Table 1 shows that when the disturbances are sampled from a normal distribution, as expected, the mismeasured regressor coefficient is downward biased for the FE case, but the

Table 1: Location Shift Model: Bias and RMSE of estimators for normal distribution ($T = 10$ and $N = 100$)

		FE	IVFE	QRFE	IVQRFE
$\beta = 1$	Bias	-0.196	0.006	-0.197	0.010
	RMSE	0.1967	0.0572	0.1978	0.0736
$\alpha = 1$	Bias	0.002	0.001	0.003	0.001
	RMSE	0.0362	0.0462	0.0447	0.0475

Table 2: Location Shift Model: Bias and RMSE of estimators for t_3 distribution ($T = 10$ and $N = 100$)

		FE	IVFE	QRFE	IVQRFE
$\beta = 1$	Bias	-0.201	-0.004	-0.159	0.009
	RMSE	0.2075	0.0619	0.1531	0.0542
$\alpha = 1$	Bias	0.001	0.001	-0.001	-0.001
	RMSE	0.0396	0.0471	0.0342	0.0363

IVFE is approximately unbiased. Similarly, in the presence of errors in variables, the QRFE estimator proposed by Koenker (2004) is downward biased and IVQRFE is able to reduce the bias. In summary, estimates are biased in both the FE and the QRFE cases, and the IV strategy is able to considerably diminish the bias for both LS and QR cases. Regarding the RMSE, in the Gaussian condition, the LS based estimators perform better than the respective QR estimators. Thus, under Gaussian conditions, the IVQRFE is capable to reduce the bias but it has a larger RMSE when compared with the IVFE.

Table 2 presents simulations for the t_3 -distribution case. The β estimates of FE and QRFE are biased, and the FE has a larger bias when compared with the same estimator in the Gaussian case. The IVQRFE is approximately an unbiased estimator for both coefficients. Interestingly, in the non-Gaussian heavy-tailed condition, the RMSE of IVQRFE is smaller than the RMSE of IVFE, evidencing that there are gains in efficiency by using a robust estimator.

Tables 3 and 4 present bias and RMSE results for the location-scale-shift model for the normal and t_3 distributions respectively. In both cases the FE and QRFE estimators are downward biased and the IVFE and IVQRFE are approximately unbiased. Note that the IVQRFE estimator presents a considerable smaller bias than QRFE and much more precision when compared with IVFE in the t_3 case.

Table 3: Location-Scale Shift Model: Bias and RMSE of estimators for normal distribution ($T = 10$ and $N = 100$)

		FE	IVFE	QRFE	IVQRFE
$\beta = 1$	Bias	-0.199	0.005	-0.195	0.009
	RMSE	0.1970	0.0569	0.1962	0.0685
$\alpha = 1$	Bias	0.001	0.001	0.000	0.001
	RMSE	0.0533	0.0579	0.0566	0.0622

Table 4: Location-Scale Shift Model: Bias and RMSE of estimators for t_3 distribution ($T = 10$ and $N = 100$)

		FE	IVFE	QRFE	IVQRFE
$\beta = 1$	Bias	-0.208	0.004	-0.146	0.007
	RMSE	0.2086	0.0779	0.1475	0.0480
$\alpha = 1$	Bias	0.002	0.003	0.004	0.001
	RMSE	0.2797	0.2801	0.0791	0.0789

4 Empirical application: Investment, Tobin's q and cash flow

As pointed out in Wansbeek (2001), the most frequent application of measurement error techniques is in the estimation of investment models using Tobin's q . This variable is the ratio of the market valuation of a firm and the replacement value of its assets. Firms with a high value of q are considered attractive as to the investment opportunities, whereas a low value of q indicates the opposite. Since the operationalization of q is not clear-cut and unambiguous, estimation poses a measurement error problem. Many empirical investment studies found a very disappointing performance of the q theory of investment, although this theory has a good performance when measurement error is purged as in Erickson and Whited (2000). In parallel, investment theory is also interested in the effect of cash flow, as the theory predicts that financially constrained firms are more likely to rely on internal funds to finance investment (see for instance Altı, 2003; Almeida, Campello and Weisbach, 2004). However, Erickson and Whited (2000) argue that cash flow has no effect on investment once measurement error in Tobin's q is taken into account.

There are distinct advantages in exploiting panel data on individual firms (see for instance Blundell, Bond, Devereux and Schiantarelli, 1992). In the first place it allows testing the theory, developed in the context of a 'representative' firm, in the context of many firms,

so reducing econometric problems introduced by aggregation across firms. Aggregation problems may be important here since the standard q model is specified in ratios and there are clear nonlinearities. Secondly, the estimates are obtained by using both the time-series and cross-sectional variation in the data. This contributes to their precision and also allows consistent estimation in the presence of firm specific effects correlated both with investment and q values. Although not reported, pooled and FE estimates differ in our estimation, which determines that firm individual FE need to be controlled for in our model.

We argue that QR is useful to describe differences in investment ratios across firms in the presence of unobserved heterogeneity. Unobserved characteristics (for the econometrician) across firms may be due to several reasons. The ability of managers (insiders) or the characteristics of investors (outsiders) can be named as the main unobservable component. For instance, Bertrand and Schoar (2003) find that a significant extent of the heterogeneity in investment, financial, and organizational practices of firms can be explained by managers 'style'. Walentin and Lorenzoni (2007) develop a model of financially constrained firms financed partly by insiders, who control its assets, and partly by outside investors. When their wealth is scarce, insiders earn a rate of return higher than the market rate of return, that is, they receive a quasi-rent on invested capital. Therefore, differences in insiders and outsiders characteristics across firms may contribute to differences in investment behavior. Additional heterogeneity may be due to differences in firms' capital structure. Chirinko (1993) shows that when the q theory is expanded for the possibility that the value of the firm depends on two or more capital inputs with differing adjustment cost technologies, the econometric equation following from optimizing behavior includes q as well as additional variables (e.g. inventory, research and development). In the upper conditional quantiles of investment, investment may be driven by insiders knowledge of business opportunities or a particular capital structure that requires more investment, and therefore, we expect that investment would be more responsive to changes in q and cash flow, as the firms would use all available resources to finance its projects. Moreover, high investment ratios reduce the marginal productivity of additional investment, and therefore, it would be difficult to get additional resources. In other words, q and cash flow sensitivities will be higher in the upper quantiles of investment than in the lower quantiles.

In contrast with the interpretation of QR in cross-section, by controlling for the firm-specific characteristics, the quantiles should be interpreted in terms of the pairs firm-year. In this case, low (high) quantiles refer to firms in periods with unusually low (high) investment

ratios when conditioning on other time-variant observable information. This determines that the same firm can be at different quantiles in different periods, which is in line with the idea that firm’s behavior will change depending on their investment needs.

The baseline model in the literature is

$$I_{it}/K_{it} = \eta_i + \beta q_{it}^* + \alpha CF_{it}/K_{it} + u_{it}, \quad (26)$$

where I denotes investment, K capital stock, CF cash flow, q^* well-measured Tobin’s q , η is the firm-specific effects and u is the innovation term. Our objective is estimating the following conditional quantile function:

$$Q_{I/K}(\tau|\eta, CF/K, q) = \eta_i(\tau) + \beta(\tau)q_{it}^* + \alpha(\tau)CF_{it}/K_{it}. \quad (27)$$

As mentioned earlier, if q is subject to measurement error, estimates of $\beta(\tau)$ will be biased down (assuming $\beta(\cdot) > 0$), but this problem can be ameliorated by using instrumental variables.

We follow Almeida, Campello, and Weisbach (2004) approach by considering a sample of manufacturing firms (SICs 2000 to 3999) over the 1980 to 2005 period with data available from COMPUSTAT’s P/S/T, Full Coverage. Only firms with observations in every year are used, in order to construct a balanced panel of firms for the 26 year period.¹³ Moreover, following those authors we eliminate firms for which cash-holdings exceeded the value of total assets and those displaying asset or sales growth exceeding 100%. Our final sample consists of 4550 firm-years and 175 firms. Summary statistics of the variables employed in our econometric specifications appear in Table 5. Because we only consider firms that report information in each of the 26 years, the sample consists mainly of relatively large firms. Although this reduces the sample heterogeneity in terms of credit constraints (e.g. presumably firms with market information for continuous 26 years are credit unconstrained), it serves for our purposes of studying firm heterogeneity along different quantiles.

Table 5: Summary statistics

Variable	Mean	Std.Dev.	0.25 quantile	Median	0.75 quantile
I/K	0.203	0.145	0.119	0.174	0.249
q	0.778	0.265	0.617	0.738	0.873
CF/K	0.414	0.444	0.211	0.345	0.534

¹³The choice of a balanced panel is made to reduce the computational burden.

$\Delta I/K$	-0.003	0.168	-0.050	-0.000	0.045
Δq	-0.004	0.136	-0.057	-0.004	0.049
$\Delta CF/K$	-0.000	0.302	-0.066	0.006	0.074

Table 6 report the regression estimates. The first column reports least squares within group (FE) estimators without (first set of rows) and with instruments (last set of rows, IVFE). The second column reports least squares in first differences for the same set of rows (FD and IVFD respectively). Following Biorn (2000) we use lags of the first differences of the mismeasured variable as instruments for fixed effects, and lags in levels for the first differences estimator. The last five columns report QR estimates with fixed effects (QRFE). Our proposed estimator, IVQRFE, appears in the last rows where the same instruments used in FE are employed. IV estimates are reported with one instrument (Δq_{t-1} for the model in levels, q_{t-2} for FD) and with two instruments ($\Delta q_{t-1}, \Delta q_{t-2}$ for the model in levels, q_{t-2}, q_{t-3} for FD) to assess the sensitivity of the estimates to the instrument selection.¹⁴

Table 6: Econometric results

			$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Variable	FE	FD	QRFE				
q	0.087 (0.010)	0.100 (0.018)	0.043 (0.009)	0.050 (0.009)	0.078 (0.009)	0.089 (0.013)	0.099 (0.017)
CF/K	0.095 (0.006)	0.053 (0.008)	0.049 (0.007)	0.060 (0.009)	0.088 (0.009)	0.112 (0.013)	0.142 (0.014)
	IVFE	IVFD	IVQRFE				
q	0.282 (0.074)	0.224 (0.081)	0.092 (0.033)	0.144 (0.045)	0.275 (0.089)	0.211 (0.104)	0.426 (0.219)
CF/K	0.074 (0.010)	0.053 (0.008)	0.043 (0.008)	0.052 (0.008)	0.066 (0.012)	0.111 (0.025)	0.123 (0.032)
Inst.	Δq_{t-1}	q_{t-2}	Δq_{t-1}				
q	0.203 (0.051)	0.181 (0.080)	0.086 (0.022)	0.116 (0.033)	0.245 (0.057)	0.223 (0.082)	0.322 (0.167)
CF/K	0.083 (0.008)	0.053 (0.008)	0.044 (0.029)	0.056 (0.029)	0.067 (0.046)	0.109 (0.061)	0.132 (0.078)
Inst.	$\Delta q_{t-1}, \Delta q_{t-2}$	q_{t-2}, q_{t-3}	$\Delta q_{t-1}, \Delta q_{t-2}$				

Notes: Standard errors in parenthesis. Sample of 4550 firm-years and 175 firms. FE: fixed effects least squares; FD: first differences least squares; QRFE: quantile regression with fixed effects; IVFE: instrumental

¹⁴Although not reported, we also use least squares projections as a different set of instruments in the QR estimations. We consider the projection of q_{it} on each respective instrument and the exogenous variables, i.e. the fixed effects and cash flow. In every case, we obtain similar results.

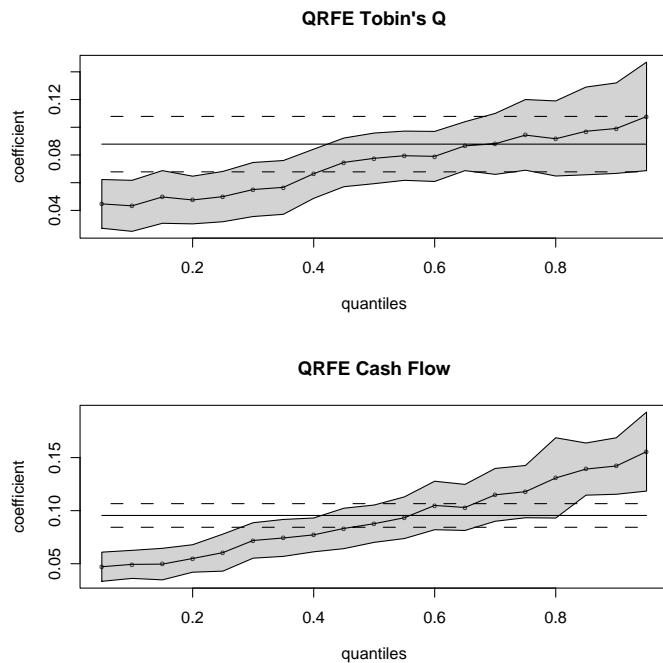
variables fixed effects least squares; IVFD: instrumental variables fixed differences least squares; IVQRFE: instrumental variables quantile regression fixed effects.

Both FE and FD estimators give similar results for the effect of Tobin's q on investment, implying that firms with higher q values would invest more. In particular, the q -sensitivity of investment is close to 0.10. Cash flow appear as positive and statistically significant. However, the FD estimate of cash flow is half of that in FE. When instrumental variables are used, the effect of Tobin's q raises considerably, which is in line with the presence of measurement error in this variable. By contrasting the effect of the instrumental variables across different estimates, it can be concluded that measurement error in Tobin's q reduces its effect by a half. Moreover, the point estimates of cash flow do not change, which can be interpreted as the fact that measurement error in Tobin's q has no spillover measurement error to cash flow. In turn, this contradicts Erickson and Whited (2000) statement that cash flow "does not matter" as it has a highly significant (mean) effect.

FE, FD and QRFE point estimates differ both in terms of Tobin's q and cash flow effects. QRFE estimates are monotonically increasing in τ , ranging from 0.045 for $\tau = 0.05$ to 0.11 at $\tau = 0.95$ (see figure 1). In terms of the Tobin's q point estimates, QRFE is below FE for $\tau < 0.75$ and below FD almost everywhere. As a result, firms-years with unusually high investment ratios will be very responsive to changes in their market valuation. In a similar token, the effect of cash flow on investment is monotonically increasing in τ . This determines that firms with conditionally high investment ratios have a higher propensity to use additional internal resources to finance investment than those with low investment ratios.

The use of IVQRFE differ from QRFE point estimates in a similar way as IVFE and IVFD differ from FE and FD respectively (see figures 2 and 3). However, in this case, the monotonicity is weakened in the q -sensitivity of investment. For low quantiles, QRFE and IVQRFE provide similar point estimates. However, both estimators differ greatly as τ increases. This suggests an interesting asymmetry in terms of the measurement error problem, as it may manifest itself mostly for firms-years in the upper quantiles of the investment distribution. Point estimates for IVQRFE are particularly high for $\tau \geq 0.90$, reaching a q -sensitivity of 0.4. Nevertheless, IVQRFE has bigger standard errors than both IVFE and QRFE. In fact for $\tau > 0.25$ the IVFE confidence interval is contained in the IVQRFE interval, while QRFE interval is always contained. In addition, QRFE and IVQRFE give similar estimates in terms of the cash flow sensitivity of investment, which reassures the fact that

Figure 1: FE and QRFE



the effect of this variable on investment is not due to measurement error in Tobin's q and that its effect is increasing in τ . Finally, it can be observed that the effect of expanding the instruments' set on the IVQRFE point estimates is in line with that in least squares model. That is, using only Δq_{t-1} produces a higher estimate for the q -sensitivity of investment than when Δq_{t-1} and Δq_{t-2} are used.

Figure 2: IVFE and IVQRFE (Instrument Δq_{t-1})

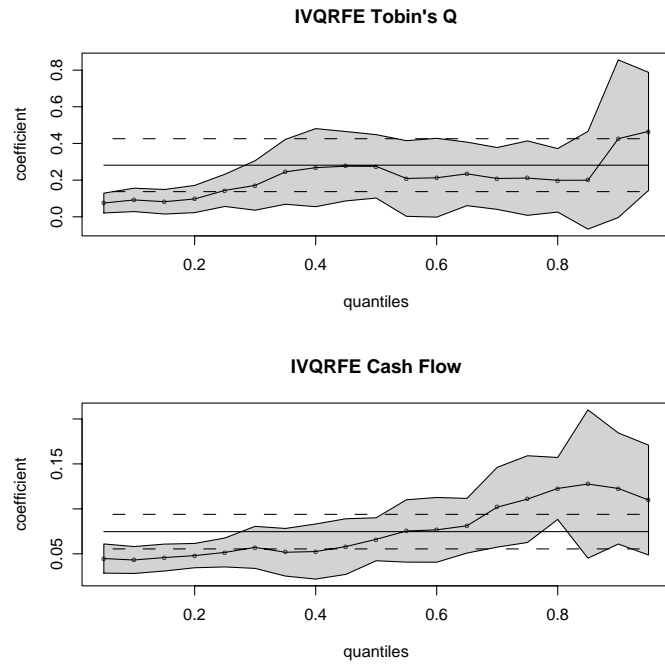
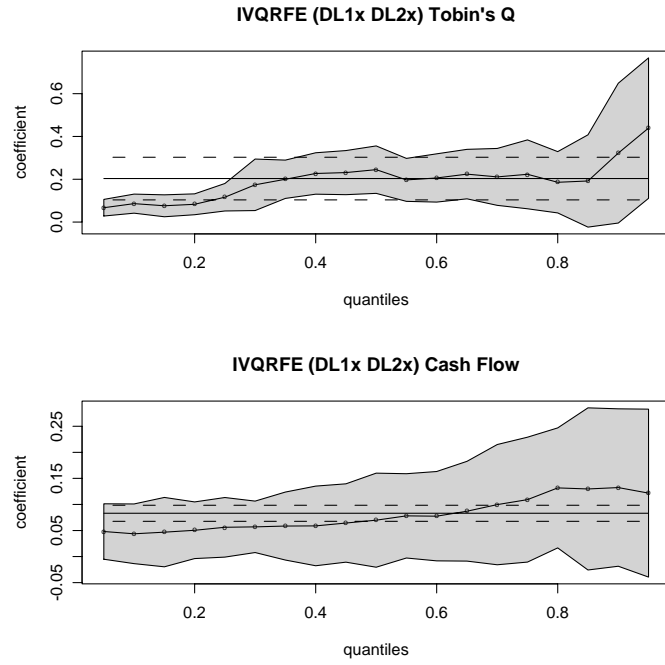


Figure 3: IVFE and IVQRFE (Instruments $\Delta q_{t-1}, \Delta q_{t-2}$)



5 Conclusion

We propose the use of an instrumental variables strategy for estimation and inference in quantile regression FE panel data models with measurement errors. Quantile regression methods allow one to explore a range of conditional quantiles exposing a variety of forms of conditional heterogeneity under less compelling distributional assumptions and provide a framework for robust estimation and inference. Most instrumental variables estimators are at least partly based on the intuition that first differencing yields a model free of FE. Despite its appeal as a procedure that avoids the estimation of the individual effects, differencing transformation to eliminate the FE is not available in the quantile regression framework, and we are required to deal directly with the full problem. We derive an approximation to the bias in the QRFE estimator in the presence of a mismeasured variable. We suggest the use of an instrumental variables strategy for estimation based on Chernozhukov and Hansen (2006, 2008) that reduces the bias in the resulting point estimates. We show that the IV FE panel estimator is consistent and asymptotically normal, provided that both N and T grow to infinity and $N^a/T \rightarrow 0$, for some $a > 0$. In addition, we suggest Wald and Kolmogorov-Smirnov type of tests for general linear hypotheses, and derive the respective limiting distributions.

Monte Carlo studies are conducted to evaluate the finite sample properties of the quantile regression instrumental variables estimator for different types of distributions. It is shown that the estimator proposed by Koenker (2004) is severely biased in the presence of mismeasured variables, while the IVQRFE sharply reduces the bias in this environment. In addition, IVQRFE has a better performance *vis-a-vis* the IV least squares based approach in terms of the bias and root mean square error of the estimators for non-Gaussian heavy-tailed distributions.

Finally, we apply the instrumental variables estimator to the most frequent application of measurement error techniques, the Tobin's q theory of investment in corporate finance. We show that the IVQRFE estimator purges for measurement error in a similar way to standard IV techniques in the least squares framework, but it shows interesting asymmetries in terms of the measurement error correction across quantiles.

Appendix 0: η 's with Pure Location Shift

The η 's are intended to capture some individual specific source of variability, or “unobserved heterogeneity,” that was not adequately controlled for by other covariates. In most applications the time series dimension T is relatively small compared to the number of individuals N . Therefore, it might be difficult to estimate a τ -dependent distributional individual effect. In this case, we still can estimate an individual specific location-shift effect.

Consider the following model for conditional quantile functions

$$Q_{y_{it}}(\tau|x_{it}^*, z_{it}) = \eta_i + x_{it}^{*'}\beta(\tau) + z_{it}'\alpha(\tau), \quad (28)$$

where $x_{it} = x_{it}^* + \epsilon_{it}$.

In this formulation the η 's have a pure location shift effect on the conditional quantiles of the response variable. The effects of the covariates (x_{it}, z_{it}) are allowed to depend upon the quantile, τ , of interest but the η 's do not. From the availability of a valid instrument, w , following Koenker (2004) and Chernozhukov and Hansen (2006, 2008) to estimate model (28) for several quantiles simultaneously we propose the following estimator

$$\hat{\beta}(\tau_k) = \underset{\beta}{\operatorname{argmin}} \|\hat{\gamma}(\beta, \tau_k)\|, \quad (29)$$

where

$$(\hat{\eta}(\beta), \hat{\alpha}(\beta, \tau), \hat{\gamma}(\beta, \tau)) = \underset{\eta, \alpha, \gamma}{\operatorname{argmin}} \sum_{k=1}^K \sum_{i=1}^N \sum_{t=1}^T v_k \rho_\tau(y_{it} - d_{it}'\eta - x_{it}'\beta(\tau_k) - z_{it}'\alpha(\tau_k) - w_{it}'\gamma(\tau_k)), \quad (30)$$

and v_k are the weights that control the relative influence of the K quantiles $\{\tau_1, \dots, \tau_K\}$ on the estimation of the η_i parameter. The design matrix for the problem of estimating $K > 1$ is as follows:

$$[v \otimes (I_n \otimes 1_T) : \Upsilon \otimes X : \Upsilon \otimes Z : \Upsilon \otimes W],$$

where I_n is a $N \times N$ identity matrix, 1_T is a $T \times 1$ vector of ones, Υ is a $K \times K$ diagonal matrix with the weights v in the diagonal. Moreover, as Koenker (2004) observes, in typical applications the above design matrix of the full problem is very sparse, i.e. has mostly zero elements. Standard sparse matrix storage schemes only require space for the non-zero elements and their indexing locations, and this considerably reduces the computational effort and memory requirements in large problems.

The optimization may become very large depending on the number of estimated quantiles. Therefore, instead of using a grid search we use a numerical optimization function in `R` (`optim`). As starting values, we use the parameters of “naive” estimation (28) without any instruments. It is important to notice that all the results presented previously are still valid. In addition, inference is essentially the same to that presented in section 2. However, the elements of the variance-covariance matrix in Theorem 2 should now be as in the design matrix presented above.

Appendix 1: Analogy between QRIV and 2SLS

Regular OLS Case

Consider the following model

$$y = X'\beta + Z'\alpha + u$$

where X is the endogenous variable, Z is the exogenous covariate and u is the error term. Let W be a valid instrument for X . In the two stage least squares procedure

$$\hat{\beta} = (X'P_{M_ZW}X)^{-1}X'P_{M_ZW}y, \quad (31)$$

where $P_{M_ZW} = M_ZW(W'M_ZW)^{-1}W'M_Z$. Note that if the same number of columns in X is the same as in W , then $W'M_ZX$ is an invertible matrix and we can simplify the above equation to

$$\hat{\beta} = (W'M_ZX)^{-1}W'M_Zy. \quad (32)$$

Grid Case

Now we consider the Grid case. Define a grid for β , $\{\beta_j, j = 1, \dots, J\}$, and for given β_j , consider the following regression

$$y - X'\beta_j = Z'\alpha + \hat{W}'\gamma + v,$$

where \hat{W} is an instrument for X , defined as the projection of X on Z and W ,

$$\hat{W} := \hat{X} = Z'\hat{\lambda} + W'\hat{\delta}.$$

The estimator $\hat{\gamma}(\beta_j)$ is

$$\hat{\gamma}(\beta_j) = (\hat{W}'M_Z\hat{W})^{-1}\hat{W}'M_Z(y - X\beta_j).$$

Now define the estimator of $\hat{\beta}$ as

$$\hat{\beta} = \arg \min_{\beta} \|\hat{\gamma}(\beta)\|_A,$$

where $\|\hat{\gamma}(\beta)\|_A = \hat{\gamma}(\beta)'A\hat{\gamma}(\beta)$, $\hat{\gamma}(\beta)$ is the vector containing the estimates $\hat{\gamma}(\beta_j)$, and A is the identity matrix. Noticing that $M_Z\hat{W} = M_ZW(W'M_ZW)^{-1}W'M_ZX$, and from the first order condition

$$\hat{\beta} = (X'M_ZW(W'M_ZW)^{-1}W'M_ZX)^{-1}X'M_ZW(W'M_ZW)^{-1}W'M_Zy.$$

Using the definition of P_{M_ZW} we have

$$\hat{\beta} = (X'P_{M_ZW}X)^{-1}X'P_{M_ZW}y, \quad (33)$$

which is the same estimator as in equation (31). When the model is exactly identified

$$\hat{\beta} = (W'M_ZX)^{-1}W'M_Zy, \quad (34)$$

which is the same estimator as in equation (32).

Appendix 2: Proof of the Theorems

The next three Lemmas help in the derivation of the results. The first Lemma shows the identification of the parameters. Lemma 2, guarantees a law of large numbers. The third Lemma states an *Argmax Process* argument which is helpful in the derivation of consistency. Later we show consistency of $\hat{\theta}(\tau)$.

Lemma 1 Given assumptions A1-A6, $(\eta(\tau), \beta(\tau), \alpha(\tau))$ uniquely solves the equations $E[\psi(Y - D\eta - X\beta - Z\alpha)\check{X}(\tau)] = 0$ over $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$.

Proof. We want to show that $(\eta(\tau), \beta(\tau), \alpha(\tau))$ uniquely solves the limit problem for each τ , that is $\eta(\beta^*(\tau), \tau) = \eta(\tau)$, $\beta^*(\tau) = \beta(\tau)$, and $\alpha(\beta^*(\tau), \tau) = \alpha(\tau)$. Let $\psi(u) := (\tau - I(u < 0))$. Define:

$$\begin{aligned} \Pi(\eta, \beta, \alpha, \tau) &:= E[\psi(Y - D'\eta(\tau) - X'\beta(\tau) - Z'\alpha(\tau))\check{X}(\tau)] \\ H(\eta, \beta, \alpha, \tau) &:= \frac{\partial}{\partial(\eta, \beta, \alpha)} E[\psi(Y - D'\eta(\tau) - X'\beta(\tau) - Z'\alpha(\tau))\check{X}(\tau)] \end{aligned}$$

By assumption A2 $H(\eta, \beta, \alpha, \tau)$ is continuous in (η, β, α) and full rank, uniformly over $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$. Moreover, by A2, the image of the set $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$ under the mapping

$(\eta, \beta, \alpha) \rightarrow \Pi(\eta, \beta, \alpha, \tau)$ is assumed to be simply connected. As in Chernozhukov and Hansen (2005), the application of Hadamard's global univalence theorem for general metric spaces, e.g. Theorem 1.8 in Ambrosetti and Prodi (1995), yields that the mapping $\Pi(\cdot, \cdot, \tau)$ is a homeomorphism (one-to-one) between $(\mathcal{E} \times \mathcal{B} \times \mathcal{A})$ and $\Pi(\mathcal{E}, \mathcal{B}, \mathcal{A}, \tau)$, the image of $\mathcal{E} \times \mathcal{B} \times \mathcal{A}$ under $\Pi(\cdot, \cdot, \tau)$. Since $(\eta, \beta, \alpha) = (\eta(\tau), \beta(\tau), \alpha(\tau))$ solves the equation $\Pi(\eta, \beta, \alpha, \tau) = 0$; and thus it is the only solution in $(\mathcal{E} \times \mathcal{B} \times \mathcal{A})$. This argument applies for every $\tau \in \mathcal{T}$. So, we have that the true parameters $(\eta, \beta, \alpha) = (\eta(\tau), \beta(\tau), \alpha(\tau))$ uniquely solve the equation

$$E[\psi(Y - D'\eta - X'\beta - Z'\alpha - W'0)\check{X}(\tau)] = 0. \quad (35)$$

Define $\phi := (\eta, \alpha, \gamma)$. By assumption and in view of the global convexity of $Q(\beta, \phi, \tau)$ in ϕ for each τ and β , $\phi(\beta, \tau)$ is defined by the subgradient condition

$$E[\psi(Y - D'\eta(\beta, \tau) - X'\beta - Z'\alpha(\beta, \tau) - W'\gamma(\beta, \tau))\check{X}(\tau)]v \geq 0 \quad (36)$$

for all $v : \phi(\beta, \tau) + v \in \mathcal{E} \times \mathcal{A} \times \mathcal{G}$.

In fact, if $\phi(\beta, \tau)$ is in the interior of $\mathcal{E} \times \mathcal{A} \times \mathcal{G}$, it uniquely solves the first order condition version of (36)

$$E[\psi(Y - D'\eta(\beta, \tau) - X'\beta - Z'\alpha(\beta, \tau) - W'\gamma(\beta, \tau))\check{X}(\tau)] = 0. \quad (37)$$

We need to find $\beta^*(\tau)$ by minimizing $\|\gamma(\beta, \tau)\|$ over β subject to (36) holding. By (35) $\beta^*(\tau) = \beta(\tau)$ makes $\|\gamma(\beta^*, \tau)\| = 0$ and satisfies (37) and hence (36) at the same time. According to the preceding argument, it is the only such solution. Thus, also by (37) $(\eta(\beta^*(\tau), \tau) = \eta(\tau), \alpha(\beta^*(\tau), \tau) = \alpha(\tau))$. ■

Lemma 2

Let $\xi_{it}(\tau) = d'_{it}\eta(\tau) + x'_{it}\beta(\tau) + z'_{it}\alpha(\tau) + w'_{it}\gamma(\tau)$ and $u_{it}(\tau) = y_{it} - \xi_{it}(\tau)$. Let $\vartheta := (\eta, \beta, \alpha, \gamma)$ be a parameter vector in $\mathcal{V} := \mathcal{E} \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$. Let

$$\delta = \begin{pmatrix} \delta_\eta \\ \delta_\beta \\ \delta_\alpha \\ \delta_\gamma \end{pmatrix} = \begin{pmatrix} \sqrt{T}(\eta - \eta(\tau)) \\ \sqrt{NT}(\beta - \beta(\tau)) \\ \sqrt{NT}(\alpha - \alpha(\tau)) \\ \sqrt{NT}(\gamma - \gamma(\tau)) \end{pmatrix}.$$

Under conditions A1-A6,

$$\begin{aligned} & \sup_{\vartheta \in \mathcal{V}} (NT)^{-1} \left| \sum_i \sum_t \left[\rho_\tau(u_{it}(\tau) - \frac{d'_{it}\delta_\eta}{\sqrt{T}} - \frac{x'_{it}\delta_\beta}{\sqrt{NT}} - \frac{z'_{it}\delta_\alpha}{\sqrt{NT}} - \frac{w'_{it}\delta_\gamma}{\sqrt{NT}}) - \rho_\tau(u_{it}(\tau)) \right. \right. \\ & \left. \left. - E\left[\rho_\tau(u_{it}(\tau) - \frac{d'_{it}\delta_\eta}{\sqrt{T}} - \frac{x'_{it}\delta_\beta}{\sqrt{NT}} - \frac{z'_{it}\delta_\alpha}{\sqrt{NT}} - \frac{w'_{it}\delta_\gamma}{\sqrt{NT}}) - \rho_\tau(u_{it}(\tau)) \right] \right| = o_p(1). \end{aligned} \quad (38)$$

Proof. With some abuse of notation let $\tilde{x}_{it} = (x_{it}, z_{it}, w_{it})$, $\tilde{\beta} = (\beta, \alpha, \gamma)$. Let $\|\cdot\|$ denote the Euclidean norm.

It is sufficient to show that for any $\epsilon > 0$

$$P\left(\sup_{\vartheta \in \mathcal{V}} (NT)^{-1} \left| \sum_i \sum_t \left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau)) \right. \right. \right. \\ \left. \left. \left. - E\left[\rho(u_{it} - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho(u_{it})\right] \right| \right| > \epsilon \right) \rightarrow 0. \quad (39)$$

Partition the parameter space $\Gamma = \mathcal{V}$ into K_N disjoint parts $\Gamma_1, \Gamma_2, \dots, \Gamma_{K_N}$, such that the diameter of each part is less than $q_{NT} = \frac{\epsilon N^\alpha}{12C_1 T}$. Let p_N be the dimension of ϑ , $p_N = O(N)$, and then $K_N \leq (2\sqrt{p_N}/q_{NT} + 1)^{p_N}$ (c.f. Wei and He (2006)).

Let $\zeta_i \in \Gamma_i, i = 1, \dots, K_N$ be fixed points. Then the left-hand side of (39) can be bounded by $P_1 + P_2$, where

$$P_1 = P\left(\max_{1 \leq q \leq K_N} \sup_{\vartheta \in \Gamma_q} (NT)^{-1} \left| \sum_i \sum_t \left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau)) \right. \right. \right. \\ \left. \left. \left. - \rho(u_{it} - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) + \rho_\tau(u_{it}(\tau)) \right. \right. \right. \\ \left. \left. \left. - E\left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))\right] \right. \right. \right. \\ \left. \left. \left. + E\left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))\right] \right| \right| > \epsilon/2 \right)$$

and

$$P_2 = P\left(\max_{1 \leq q \leq K_N} (NT)^{-1} \left| \sum_i \sum_t \left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) + \rho_\tau(u_{it}(\tau)) \right. \right. \right. \\ \left. \left. \left. - E\left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))\right] \right| \right| > \epsilon/2 \right).$$

Therefore, it suffices to show that bot P_1 and P_2 are $o_p(1)$. Noting that $|\rho_\tau(x + y) - \rho_\tau(x)| \leq 2|y|$, we have

$$\begin{aligned}
& \max_{1 \leq q \leq K_N} \sup_{\vartheta \in \Gamma_q} (NT)^{-1} \left| \sum_i \sum_t \left[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau)) \right. \right. \\
& \quad \left. \left. - \rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{q\zeta_{q\eta}}{\sqrt{T}}) + \rho_\tau(u_{it}(\tau)) \right. \right. \\
& \quad \left. \left. - E[\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))] \right. \right. \\
& \quad \left. \left. + E[\rho(u_{it} - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho(u_{it})] \right] \right| \\
& \leq \max_{1 \leq q \leq K_N} \sup_{\vartheta \in \Gamma_q} (NT)^{-1} \left[2 \left| \sum_i \sum_t \left[-\frac{\tilde{x}'_{it}}{\sqrt{NT}} (\delta_{\tilde{\beta}} - \zeta_{q\tilde{\beta}}) - \frac{d'_{it}}{\sqrt{T}} (\delta_\eta - \zeta_{q\eta}) \right] \right| \right. \\
& \quad \left. + 2 \left| \sum_i \sum_t E \left[\frac{\tilde{x}'_{it}}{\sqrt{NT}} (\delta_{\tilde{\beta}} - \zeta_{q\tilde{\beta}}) - \frac{d'_{it}}{\sqrt{T}} (\delta_\eta - \zeta_{q\eta}) \right] \right| \right] \\
& \leq \max_{1 \leq q \leq K_N} \sup_{\vartheta \in \Gamma_q} (NT)^{-1} \left[2 \sum_i \sum_t \left[\frac{1}{\sqrt{NT}} \|\tilde{x}'_{it}\| \|(\delta_{\tilde{\beta}} - \zeta_{q\tilde{\beta}})\| - \frac{1}{\sqrt{T}} \|d'_{it}\| \|(\delta_\eta - \zeta_{q\eta})\| \right] \right. \\
& \quad \left. + 2 \sum_i \sum_t E \left[\frac{1}{\sqrt{NT}} \|\tilde{x}'_{it}\| \|(\delta_{\tilde{\beta}} - \zeta_{q\tilde{\beta}})\| - \frac{1}{\sqrt{T}} \|d'_{it}\| \|(\delta_\eta - \zeta_{q\eta})\| \right] \right] \\
& \leq \left[4q_{NT} \right] \leq \frac{\epsilon}{2},
\end{aligned}$$

where the last equality follows by assumptions A5-A6, and implies $P_1 = o_p(1)$. Thus, it remains to show that $P_2 = o_p(1)$.

If we write $m_{itq} = [\rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))]$ en

$$\begin{aligned}
P_2 &= P \left(\max_{1 \leq q \leq K_N} (NT)^{-1} \left| \sum_i \sum_t [m_{itq} - E m_{itq}] \right| > \epsilon/2 \right) \\
&\leq \sum_{q=1}^{K_N} P \left((NT)^{-1} \left| \sum_i \left[\sum_t m_{itq} - E \sum_t m_{itq} \right] \right| > \epsilon/2 \right)
\end{aligned}$$

For fixed i ,

$$\begin{aligned}
\sum_t m_{itq} &\leq \sum_t \sup_{\zeta \in \Gamma_q} \left| \rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\zeta_{\tilde{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\zeta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau)) \right| \\
&\leq \sum_t \sup_{\zeta \in \Gamma_q} 2 \left| \tilde{x}'_{it} \frac{\zeta_{\tilde{\beta}}}{\sqrt{NT}} + d'_{it} \frac{\zeta_\eta}{\sqrt{T}} \right| \\
&\leq \sum_t \sup_{\zeta \in \Gamma_q} 2 \left[\frac{1}{\sqrt{NT}} \|\tilde{x}'_{it}\| \|\zeta_{\tilde{\beta}}\| + \frac{1}{\sqrt{T}} \|d'_{it}\| \|\zeta_\eta\| \right] \leq TC_2,
\end{aligned}$$

and,

$$\text{Var}\left(\sum_{t=1}^T m_{itq}\right) \leq T \sum_{t=1}^T \text{Var}(m_{itq}).$$

Note that,

$$\left|\tilde{x}_{it} \frac{\zeta_{q\tilde{\beta}}}{\sqrt{NT}} + d'_{it} \frac{\zeta_{q\eta}}{\sqrt{T}}\right|^2 \leq C_3 \|\zeta_{q\tilde{\beta}}\|^2 + C_4 \|\zeta_{q\eta}\|^2$$

Using the fact that $\|\zeta_q\| < C$ for any q and the independence across individuals in A1, we have

$$\sum_i \text{Var} \sum_t m_{itq} \leq T \sum_i \sum_t E m_{itq}^2 \leq NT^2 C_5 + o(1)$$

By the assumption of independence, we can bound P_2 using Bernstein's inequality:

$$\begin{aligned} P_2 &\leq 2K_N \exp\left(-\frac{\left(\frac{\epsilon NT}{2}\right)^2}{\sum_i \text{Var} \sum_t m_{it} + \frac{TC_2}{3} \left(\frac{\epsilon NT}{2}\right)}\right) \\ &\leq 2(2\sqrt{p_N}/q_{NT} + 1)^{p_N} \exp\left(-\frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2 C_5 + o(1) + \frac{TC_2}{3} \left(\frac{\epsilon NT}{2}\right)}\right) \\ &= 2 \exp\left(p_N \ln(2\sqrt{p_N}/q_{NT} + 1) - \frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2 C_5 + o(1) + \frac{TC_2}{3} \left(\frac{\epsilon NT}{2}\right)}\right) \\ &= 2 \exp\left(O(N) \ln(24O(N^{1/2-a}T)C_1/\epsilon + 1) - \frac{\left(\frac{\epsilon NT}{2}\right)^2}{NT^2 C_5 + o(1) + \frac{TC_2}{3} \left(\frac{\epsilon NT}{2}\right)}\right) \end{aligned}$$

Hence, $P_2 \rightarrow 0$ as $N, T \rightarrow \infty$ and $N^a/T \rightarrow 0$ and the lemma follows. ■

Lemma 3 (Corollary 3.2.3 - van der Vaart and Wellner (1996))

Let M_n be stochastic processes indexed by a metric space Θ , and let $M : \Theta \rightarrow \mathfrak{R}$ be a deterministic function. Suppose that $\|M_n - M\|_{\Theta} \xrightarrow{p} 0$ and that there exists a point θ_0 such that

$$M(\theta_0) > \sup_{\theta \notin D} M(\theta)$$

for every open set D that contains θ_0 . Then any sequence $\hat{\theta}_n$, such that $Q_n(\hat{\theta}_n) > \sup_{\theta} Q_n(\theta) - o_p(1)$, satisfies $\hat{\theta}_n \xrightarrow{p} \theta_0$.

Theorem 1 (Consistency):

Under conditions A1-A6

$$\hat{\varphi}(\tau) = \varphi(\tau) + o_p(1).$$

Proof.

Define

$$\mathcal{P} = (\eta, \beta, \alpha, \gamma) \rightarrow \rho_\tau(Y - D'\eta - X'\beta - Z'\alpha - W'\gamma)$$

and note that \mathcal{P} is continuous. Therefore, uniform convergence is given by Lemma 2, such that

$$\sup_{\vartheta \in \mathcal{V}} |M_n(\vartheta) - M(\vartheta)| = o_p(1)$$

where $M_n \equiv \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(u_{it}(\tau) - \tilde{x}'_{it} \frac{\delta_{\hat{\beta}}}{\sqrt{NT}} - d'_{it} \frac{\delta_\eta}{\sqrt{T}}) - \rho_\tau(u_{it}(\tau))$, and $M = EM_n$.

Therefore, denoting $\phi := (\eta', \alpha', \gamma)'$, we have that $\|\hat{\phi}(\beta, \tau) - \phi(\beta, \tau)\| \xrightarrow{p} 0$ (*) which implies that $\|\hat{\gamma}(\beta, \tau) - \gamma(\beta, \tau)\| \xrightarrow{p} 0$, which by a simple Argmax process over a compact set argument (assumption A4) implies that $\|\hat{\beta}(\tau) - \beta(\tau)\| \xrightarrow{p} 0$, which by (*) implies that $\|\hat{\alpha}(\tau) - \alpha(\tau)\| \xrightarrow{p} 0$ and $\|\hat{\gamma}(\hat{\beta}, \tau) - 0\| \xrightarrow{p} 0$. Therefore, $\|\hat{\theta}(\tau) - \theta(\tau)\| \xrightarrow{p} 0$ and the theorem follows.

■

Now we prove Theorem 2, i.e., under conditions A1-A6 the proposed estimator is asymptotically normal.

Proof.

Consider the following model

$$y_{it} = d'_{it}\eta + x'_{it}\beta + z'_{it}\alpha + u_{it}. \quad (40)$$

The objective function is

$$\min_{\eta, \beta, \alpha, \gamma} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - d'_{it}\eta - x'_{it}\beta - z'_{it}\alpha - w'_{it}\gamma). \quad (41)$$

Consider a collection of closed balls $B_n(\beta(\tau))$, centered at $\beta(\tau)$, with radius π_n , and $\pi_n \rightarrow 0$ slowly enough. Note that, for any $\beta_n(\tau) \xrightarrow{p} \beta(\tau)$ we can write the objective function as

$$V_{NT}(\delta_N) = \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \xi_{it}(\tau) - d_{it}\delta_\eta/\sqrt{T} - x'_{it}\delta_\beta/\sqrt{NT} - z'_{it}\delta_\alpha/\sqrt{NT} - w'_{it}\delta_\gamma/\sqrt{NT}) - \rho_\tau(y_{it} - \xi_{it}(\tau)) \quad (42)$$

where $\xi_{it}(\tau) = d'_{it}\eta(\tau) + x'_{it}\beta(\tau) + z'_{it}\alpha(\tau) + w'_{it}\gamma(\tau)$, and in the minimum

$$\hat{\delta}_n = \begin{pmatrix} \hat{\delta}_\eta(\beta_n(\tau)) \\ \hat{\delta}_\beta(\beta_n(\tau)) \\ \hat{\delta}_\alpha(\beta_n(\tau)) \\ \hat{\delta}_\gamma(\beta_n(\tau)) \end{pmatrix} = \begin{pmatrix} \sqrt{T}(\hat{\eta}(\beta_n(\tau), \tau) - \eta(\tau)) \\ \sqrt{NT}(\beta_n(\tau) - \beta(\tau)) \\ \sqrt{NT}(\hat{\alpha}(\beta_n(\tau), \tau) - \alpha(\tau)) \\ \sqrt{NT}(\hat{\gamma}(\beta_n(\tau), \tau) - 0) \end{pmatrix}.$$

Note that, for fixed $(\delta_\beta, \delta_\alpha, \delta_\gamma)$, we can consider the behavior of δ_η . For each i ,

$$g_i(\delta_{\eta_i}, \delta_\beta, \delta_\alpha, \delta_\gamma) = -\frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau) - \frac{\delta_{\eta_i}}{\sqrt{T}} - x'_{it} \frac{\delta_\beta}{\sqrt{NT}} - z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} - w'_{it} \frac{\delta_\gamma}{\sqrt{NT}}).$$

For, fixed $\delta_\alpha, \delta_\gamma$, $\sup_{\tau \in \mathcal{T}} \|\beta_n(\tau) - \beta(\tau)\| \xrightarrow{p} 0$, and $K > 0$

$$\sup_{|\delta_{\eta_i}| < K} \|g_i(\delta_{\eta_i}, \delta_\beta, \delta_\alpha, \delta_\gamma) - g_i(0, 0, 0, 0) - E[g_i(\delta_{\eta_i}, \delta_\beta, \delta_\alpha, \delta_\gamma) - g_i(0, 0, 0, 0)]\| = o_p(1).$$

Expanding we have

$$\begin{aligned} & E[g_i(\delta_{\eta_i}, \delta_\beta, \delta_\alpha, \delta_\gamma) - g_i(0, 0, 0, 0)] \\ &= E\left(-\frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau) - \frac{\delta_{\eta_i}}{\sqrt{T}} - x'_{it} \frac{\delta_\beta}{\sqrt{NT}} - z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} - w'_{it} \frac{\delta_\gamma}{\sqrt{NT}}) + \frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau))\right) \\ &= -\frac{1}{\sqrt{T}} \sum_{t=1}^T \left[\tau - F(\xi_{it}(\tau) + \frac{\delta_{\eta_i}}{\sqrt{T}} + x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}}) \right] \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T \left[F(\xi_{it}(\tau) + \frac{\delta_{\eta_i}}{\sqrt{T}} + x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}}) - \tau \right] \\ &= \frac{1}{\sqrt{T}} \left[\sum_{t=1}^T f_{it}(\xi_{it}(\tau)) \left[\frac{\delta_{\eta_i}}{\sqrt{T}} + x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} w'_{it} + \frac{\delta_\gamma}{\sqrt{NT}} \right] + R_{it} \right] \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^T f_{it}(\xi_{it}(\tau)) \left[\frac{\delta_{\eta_i}}{\sqrt{T}} + x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}} \right] + R_{it}. \end{aligned}$$

Optimality of $\hat{\delta}_{\eta_i}$ implies that $g_i(\delta_{\eta_i}, \delta_\beta, \delta_\alpha, \delta_\gamma) = o(T^{-1})$, and thus

$$\hat{\delta}_{\eta_i} = -\bar{f}_i^{-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T f_{it}(\xi_{it}(\tau)) \left[x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w_{it} \frac{\delta_\gamma}{\sqrt{NT}} \right] - \frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau)) \right] + R_{it}$$

where $\bar{f}_i = T^{-1} \sum_t f_{it}$ and R_{it} is the remainder term for each i .

Substituting $\hat{\delta}_\eta$'s, we denote

$$G(\delta_\beta, \delta_\alpha, \delta_\gamma) = -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{it} \psi_\tau(y_{it} - \xi_{it}(\tau) - d'_{it} \frac{\hat{\delta}_\eta}{\sqrt{T}} - x'_{it} \frac{\delta_\beta}{\sqrt{NT}} - z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} - w_{it} \frac{\delta_\gamma}{\sqrt{NT}})$$

where $\tilde{X}_{it} = (w'_{it}, z'_{it})'$, and $\delta_\beta(\beta_n) = \sqrt{NT}(\beta_n(\tau) - \beta(\tau))$, $\delta_\alpha(\beta_n) = \sqrt{NT}(\hat{\alpha}(\beta_n, \tau) - \alpha)$, and $\delta_\gamma(\beta_n) = \sqrt{NT}(\hat{\gamma}(\beta_n, \tau) - 0)$.

According to Lemma B2 in Chernozhukov and Hansen (2006)

$$\sup_{\|(\delta_\beta, \delta_\alpha, \delta_\gamma)\| < K} \|G_i(\delta_\beta, \delta_\alpha, \delta_\gamma) - G_i(0, 0, 0) - E[G_i(\delta_\beta, \delta_\alpha, \delta_\gamma) - G_i(0, 0, 0)]\| = o_p(1),$$

and at the minimizer, $G(\hat{\delta}_\beta, \hat{\delta}_\alpha, \hat{\delta}_\gamma) = o((NT)^{-1})$.

Expanding, as above,

$$\begin{aligned} & E[G(\delta_\beta, \delta_\alpha, \delta_\gamma) - G(0, 0, 0)] \\ &= \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{X}_{it} f_{it} \left(x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}} + \frac{\hat{\delta}_\eta}{\sqrt{T}} \right) + o_p(1) \\ &= \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{X}_{it} f_{it} \left[x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}} - \frac{1}{\sqrt{T}} \bar{f}_i^{-1} \left[T^{-1/2} \sum_t f_{it}(\xi_{it}(\tau)) \right. \right. \\ & \left. \left. \left[x'_{it} \frac{\delta_\beta}{\sqrt{NT}} + z'_{it} \frac{\delta_\alpha}{\sqrt{NT}} + w'_{it} \frac{\delta_\gamma}{\sqrt{NT}} \right] - T^{-1/2} \sum_{t=1}^T \psi_\tau(y_{it} - \xi_{it}(\tau)) \right] + R_{it} \right] + o_p(1) \\ &= \frac{1}{NT} \sum_i \sum_t \tilde{X}_{it} f_{it} \left(x_{it} - \bar{f}_i^{-1} T^{-1} \sum_t f_{it} x_{it} \right)' \delta_\beta + \frac{1}{NT} \sum_i \sum_t \tilde{X}_{it} f_{it} \left(z_{it} - \bar{f}_i^{-1} T^{-1} \sum_t f_{it} z_{it} \right)' \delta_\alpha \\ &+ \frac{1}{NT} \sum_i \sum_t \tilde{X}_{it} f_{it} \left(w_{it} - \bar{f}_i^{-1} T^{-1} \sum_t f_{it} w_{it} \right)' \delta_\gamma - \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{X}_{it} f_{it} \bar{f}_i^{-1} T^{-1/2} \sum_t \psi(y_{it} - \xi_{it}) \\ &- \frac{1}{\sqrt{NT}} \sum_i \sum_t \tilde{X}_{it} f_{it} R_{it} / \sqrt{T} + o_p(1), \end{aligned}$$

where the order of the final term is controlled by the bound on the derivative of the conditional density. It is important to note that $G(\hat{\delta}_\beta, \hat{\delta}_\alpha, \hat{\delta}_\gamma) = 0$ and then $E[G(\delta_\beta, \delta_\alpha, \delta_\gamma) - G(0, 0, 0)] = -G(0, 0, 0)$. Let $\delta_{\alpha\gamma} = (\delta_\alpha, \delta_\gamma)$ and let Ψ_τ be a NT -vector $(\psi_\tau(y_{it} - \xi_{it}(\tau)))$,

$$\begin{aligned} E(\tilde{X}' M_D \Phi(\tau) M_D X) \delta_\beta + E(\tilde{X}' M_D \Phi(\tau) M_D \tilde{X}) \delta_{\alpha\gamma} - \tilde{X}' P_D \Psi_\tau &= -\tilde{X}' \Psi_\tau + R_{nT} \\ J_{\alpha\gamma} \delta_{\alpha\gamma} &= [-(\tilde{X}' M_D \Psi_\tau - R_{nT}) - J_\beta \delta_\beta] \\ \hat{\delta}_{\alpha\gamma} &= J_{\alpha\gamma}^{-1} [-(\tilde{X}' M_D \Psi_\tau - R_{nT}) - J_\beta \delta_\beta] \end{aligned}$$

where

$$\begin{aligned}
M_D &= I - P_D \\
P_D &= D(D'\Phi(\tau)D)^{-1}D\Phi(\tau) \\
J_{\alpha\gamma} &= E(\tilde{X}'M_D\Phi(\tau))M_D\tilde{X}) \\
J_\beta &= E(\tilde{X}'M_D\Phi(\tau)M_DX)
\end{aligned}$$

and

$$R_{nT} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \tilde{X}_{it} f_{it} R_{it} / \sqrt{T} + o_p(1).$$

Let $[\bar{J}'_\alpha, \bar{J}'_\gamma]'$ be the conformable partition of $J_{\alpha\gamma}^{-1}$, then

$$\hat{\delta}_\gamma = \bar{J}_\gamma [-(\tilde{X}'M_D\Psi_\tau - R_{nT}) - J_\beta\delta_\beta] \quad (43)$$

$$\hat{\delta}_\alpha = \bar{J}_\alpha [-(\tilde{X}'M_D\Psi_\tau - R_{nT}) - J_\beta\delta_\beta]. \quad (44)$$

The remainder term R_{nT} has a dominant component that comes from the Bahadur representation of the η 's. By A1 and A5, we have for a generic constant K

$$R_{nT} = T^{-1/4} \frac{K}{\sqrt{N}} \sum_{i=1}^N R_{i0} + o_p(1).$$

The analysis of Knight (2001) shows that the summands converge in distribution, that is as $T \rightarrow \infty$ the remainder term $T^{1/4}R_{it} \xrightarrow{d} R_{i0}$, where R_{i0} are functionals of Brownian motions with finite second moment. Therefore, independence of y_{it} , and condition A6 ensure that the contribution of the remainder is negligible. Thus (43) and (34) simplify to

$$\begin{aligned}
\hat{\delta}_\gamma &= \bar{J}_\gamma [-(\tilde{X}'M_D\Psi(\tau)) - J_\beta\delta_\beta] \\
\hat{\delta}_\beta &= J_\beta [-(\tilde{X}'M_D\Psi(\tau)) - J_\beta\delta_\beta].
\end{aligned}$$

By consistency, $wp \rightarrow 1$

$$\hat{\delta}_\beta = \min_{\delta_\beta \in B_n(\beta(\tau))} \hat{\delta}_\gamma(\delta_\beta)' A \hat{\delta}_\gamma(\delta_\beta),$$

assuming that $\hat{\delta}_\gamma' A \hat{\delta}_\gamma$ is continuous in δ_β , and from the first order condition

$$\hat{\delta}_\beta = -[J'_\beta \bar{J}_\gamma A \bar{J}_\gamma J_\beta]^{-1} [J'_\beta \bar{J}_\gamma A \bar{J}_\gamma (\tilde{X}'M_D\Psi)].$$

Substituting $\hat{\delta}_\beta$ back in δ_α we obtain

$$\begin{aligned}\hat{\delta}_\alpha &= -\bar{J}_\alpha[(\tilde{X}'M_D\Psi) - J_\beta[J'_\beta\bar{J}_\gamma A\bar{J}_\gamma J_\beta]^{-1}J'_\beta\bar{J}_\gamma A\bar{J}_\gamma(\tilde{X}'M_D\Psi)] \\ &= -\bar{J}_\alpha[(I - J_\beta[J'_\beta\bar{J}_\gamma A\bar{J}_\gamma J_\beta]^{-1}J'_\beta\bar{J}_\gamma A\bar{J}_\gamma)(\tilde{X}'M_D\Psi)].\end{aligned}$$

It is also important to analyze $\hat{\delta}_\gamma$. Thus, replacing $\hat{\delta}_\beta$ in δ_γ

$$\begin{aligned}\hat{\delta}_\gamma &= -\bar{J}_\gamma[(\tilde{X}'M_D\Psi) - J_\beta[J'_\beta\bar{J}_\gamma A\bar{J}_\gamma J_\beta]^{-1}J'_\beta\bar{J}_\gamma A\bar{J}_\gamma(\tilde{X}'M_D\Psi)] \\ &= -\bar{J}_\gamma[(I - J_\beta[J'_\beta\bar{J}_\gamma A\bar{J}_\gamma J_\beta]^{-1}J'_\beta\bar{J}_\gamma A\bar{J}_\gamma)(\tilde{X}'M_D\Psi)].\end{aligned}$$

By condition A3, using the fact that $\bar{J}_\gamma J_\beta$ is invertible

$$\hat{\delta}_\gamma = 0 + O_p(1) + o_p(1).$$

Noticing that under assumptions A1-A6, $\Psi \xrightarrow{d} N(0, \omega)$, where $\omega = \tau(1 - \tau)$. Neglecting the remainder term, and using the definition of δ_β and δ_α we have

$$\sqrt{NT}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Omega), \quad \Omega = (K', L')'S(K', L')$$

where $S = \tau(1 - \tau)E[VV']$, $V = \tilde{X}'M_D$, $K = (J'_\beta H J_\beta)^{-1}J'_\beta H$, $H = \bar{J}'_\gamma A[\beta(\tau)]\bar{J}_\gamma$, $L = \bar{J}_\alpha M$, $M = I - J_\beta K$.

■

References

- ABREYAYA, J., AND C. M. DAHL (2007): “The Effects of Smoking and Prenatal Care on Birth Outcomes: Evidence from Quantile Estimation on Panel Data,” *Journal of Business and Economic Statistics*, forthcoming.
- ALMEIDA, H., M. CAMPELLO, AND M. S. WEISBACH (2004): “The Cash Flow Sensitivity of Cash,” *Journal of Finance*, 59, 1777–1804.
- ALTI, A. (2003): “How Sensitive Is Investment to Cash Flow When Financing Is Frictionless?,” *Journal of Finance*, 58, 707–722.
- ALVAREZ, J., AND M. ARELLANO (2003): “The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators,” *Econometrica*, 71, 1121–1159.

- AMBROSETTI, A., AND G. PRODI (1995): *A Primer of Nonlinear Analysis, Cambridge Studies in Advanced Mathematics, vol. 34.* Cambridge University Press, Cambridge, Massachusetts.
- ANDREWS, D. W. K. (1993): “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821–856.
- ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2006): “Quantile Regression Under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74, 539–573.
- ARELLANO, M., AND J. HAHN (2005): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments,” Mimeo, CEMFI.
- BERTRAND, M., AND A. SCHOAR (2003): “Managing With Style: The Effect Of Managers On Firm Policies,” *Quarterly Journal of Economics*, 118, 1169–1208.
- BESTER, C. A., AND C. HANSEN (2007): “A Penalty Function Approach to Bias Reduction in Non-Linear Panel Models with Fixed Effects,” *Journal of Business and Economic Statistics*, forthcoming.
- BIORN, E. (1992): “Econometrics of Panel Data with Measurement Errors,” in *Econometrics of Panel Data: Theory and Applications*, ed. by L. Matyas, and P. Sevestre, pp. 1322–1393. Kluwer.
- BIORN, E. (2000): “Panel Data with Measurement Errors: Instrumental Variables and GMM Procedures Combining Levels and Differences,” *Econometric Reviews*, 19, 391–424.
- BLUNDELL, R., S. BOND, M. DEVEREUX, AND F. SCHIANTARELLI (1992): “Investment and Tobin’s Q: Evidence from Company Panel Data,” *Journal of Econometrics*, 51, 233–257.
- CHAMBERLAIN, G. (1982): “Multivariate Regression Models for Panel Data,” *Journal of Econometrics*, 18, 5–46.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73, 245–261.

- (2006): “Instrumental Quantile Regression Inference for Structural and Treatment Effects Models,” *Journal of Econometrics*, 132, 491–525.
- (2008): “Instrumental Variable Quantile Regression: A Robust Inference Approach,” *Journal of Econometrics*, 142, 379–398.
- CHESHER, A. (2001): “Parameter Approximations for Quantile Regressions with Measurement Error,” Working Paper CWP02/01, Department of Economics, University College London.
- CHIRINKO, R. S. (1993): “Multiple Capital Inputs, Q, and Investment Spending,” *Journal of Economic Dynamics and Control*, 17, 907–928.
- DELONG, D. M. (1981): “Crossing Probabilities for a Square Root Boundary By a Bessel Process,” *Communications in Statistics, Part A - Theory and Methods*, 10, 2197–2213.
- ERICKSON, T., AND T. M. WHITED (2000): “Measurement Error and the Relationship between Investment and q,” *Journal of Political Economy*, 108, 1027–1057.
- GERACI, M., AND M. BOTTAI (2007): “Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution,” *Biostatistics*, 8, 140–154.
- GRILICHES, Z., AND J. A. HAUSMAN (1986): “Errors in Variables in Panel Data,” *Journal of Econometrics*, 31, 93–118.
- GUTENBRUNNER, C., AND J. JURECKOVA (1992): “Regression Rank Scores and Regression Quantiles,” *The Annals of Statistics*, 20, 305–330.
- HAHN, J., AND G. M. KUERSTEINER (2002): “Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When both n and T are Large,” *Econometrica*, 70, 1639–1657.
- HAHN, J., AND W. NEWEY (2004): “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models,” *Econometrica*, 72, 1295–1319.
- HE, X., AND Q.-M. SHAO (2000): “On Parameters of Increasing Dimensions,” *Journal of Multivariate Analysis*, 73, 120–135.
- HSIAO, C. (1992): “Nonlinear Latent Variables Models,” in *Econometrics of Panel Data: Theory and Applications*, ed. by L. Matyas, and P. Sevestre, pp. 242–261. Kluwer.

- (2003): *Analysis of Panel data*. Cambridge University Press, Cambridge, Massachusetts, 2d edn.
- HSIAO, C., M. H. PESARAN, AND A. K. TAHMISCIOLU (2002): “Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods,” *Journal of Econometrics*, 109, 107–150.
- HSIAO, C., AND G. TAYLOR (1991): “Some Remarks on Measurement Errors and the Identification of Panel data Models,” *Statistica Neerlandica*, 45, 187–194.
- KNIGHT, K. (2001): “Comparing Conditional Quantile Estimators: First and Second order Considerations,” Mimeo.
- KOENKER, R. (2004): “Quantile Regression for Longitudinal Data,” *Journal of Multivariate Analysis*, 91, 74–89.
- KOENKER, R., AND G. W. BASSETT (1978): “Regression Quantiles,” *Econometrica*, 46, 33–49.
- (1982a): “Robust Tests for Heteroscedasticity Based on Regression Quantile,” *Econometrica*, 50, 43–61.
- (1982b): “Tests of Linear Hypothesis and l_1 Estimation,” *Econometrica*, 50, 1577–1584.
- KOENKER, R., AND K. HALLOCK (2000): “Quantile Regression an Introduction,” Manuscript, University of Illinois at Urbana-Champaign.
- KOENKER, R., AND J. A. F. MACHADO (1999): “Godness of Fit and Related Inference Processes for Quantile Regression,” *Journal of the American Statistical Association*, 94, 1296–1310.
- KOENKER, R., AND Z. XIAO (2006): “Quantile Autoregression,” *Journal of the American Statistical Association*, 101, 980–990.
- LAMARCHE, C. (2006): “Robust Penalized Quantile Regression Estimation for Panel Data,” Mimeo.
- (2008): “Penalized Quantile Regression Estimation for a Model with Endogenous Individual Effects,” Mimeo.

- NEYMAN, J., AND E. L. SCOTT (1948): “Consistent Estimates Based on Partially Consistent Observations,” *Econometrica*, 16, 1–32.
- PORTNOY, S. (1984): “Tightness of the Sequence of Empiric cdf Processes Defined from Regression Fractiles,” in *Robust and Nonlinear Time Series Analysis*, ed. by J. Frank, W. Hardle, and D. Martin. Springer-Verlag: New York.
- (1985): “Asymptotic Behavior of M Estimators of p Regression Parameters when p^2/n is Large; II. Normal Approximation,” *The Annals of Statistics*, 13, 1403–1417.
- POWELL, J. L. (1986): “Censored Regression Quantiles,” *Journal of Econometrics*, 32, 143–155.
- SCHENNACH, S. M. (2004): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72, 33–75.
- (2008): “Quantile Regression with Mismeasured Covariates,” *Econometric Theory*, 24, 1010–1043.
- VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag Press, New York, New York.
- WALENTIN, K., AND G. LORENZONI (2007): “Financial Frictions, Investment and Tobin’s?,” NBER Working Paper Series 13092.
- WANSBEEK, T. (2001): “GMM Estimation in Panel Data Models with Measurement Error,” *Journal of Econometrics*, 104, 259–268.
- WANSBEEK, T., AND R. H. KONING (1989): “Measurement Error and Panel Data,” *Statistica Neerlandica*, 45, 85–92.
- WEI, Y., AND X. HE (2006): “Conditional Growth Charts,” *The Annals of Statistics*, 34, 2069–2097.