Quantiles via Moments*

José A. F. Machado
NOVA School of Business and Economics

J. M. C. Santos Silva
School of Economics, University of Surrey

8 March 2018

Abstract

We study the conditions under which it is possible to estimate regression quantiles by estimating conditional means. The advantage of this approach is that it allows the use of methods that are only valid in the estimation of conditional means, while still providing information on how the regressors affect the entire conditional distribution. The methods we propose are not meant to replace the well-established quantile regression estimator, but provide an additional tool that can allow the estimation of regression quantiles in settings where otherwise that would be difficult or even impossible. We consider two settings in which our approach can be particularly useful: panel data models with individual effects and models with endogenous explanatory variables. Besides presenting the estimator and establishing the regularity conditions needed for valid inference, we perform a small simulation experiment, present two simple illustrative applications, and discuss possible extensions.

JEL classification code: C21, C23, C26.

Key words: Endogeneity; Fixed effects; Linear heteroskedasticity; Location-scale model; Quantile regression.

*We are grateful to the Editor Xuming He and to two anonymous referees for constructive comments and advice that helped to substantially improve the paper. We also thank Geert Dhaene and Paulo Parente for useful comments and advice. The usual disclaimer applies. Address for correspondence:
José Machado, NOVA School of Business and Economics, Portugal. E-mail: jafm@novasbe.pt.
João Santos Silva, School of Economics, University of Surrey, UK. E-mail: jmcss@surrey.ac.uk.

1. INTRODUCTION

We study the conditions under which it is possible to estimate regression quantiles by estimating conditional means. We focus on the conditional location-scale model considered, among others, by Koenker and Bassett (1982), Gutenbrunner and Jurečková (1992), Koenker and Zhao (1994), He (1997), and Zhao (2000), and propose an estimator of the conditional quantiles obtained by combining estimates of the location and scale functions, both of which are identified by conditional expectations of appropriately defined variables.

The advantage of our approach is that it allows the use of methods that are only valid in the estimation of conditional means, such as differencing out individual effects in panel data models, while providing information on how the regressors affect the entire conditional distribution. These informational gains are perhaps the most attractive feature of quantile regression (see, e.g., the influential papers by Chamberlain, 1994, and Buchinsky, 1994) and were emphasized, for example, in the surveys by Koenker and Hallock (2001), Cade and Noon (2003), and Bassett and Koenker (2018). Besides greatly facilitating the estimation of complex models, our approach also leads to estimates of the regression quantiles that do not cross, a crucial requisite often ignored in empirical applications (see also He, 1997, and Chernozhukov, Fernández-Val, and Galichon, 2010).

Because our estimator is based on conditional means, it does not share some of the robustness properties of the seminal quantile regression estimator of Koenker and Bassett (1978), which is based on the check function. For example, our estimator requires stronger assumptions on the existence of moments than those needed for the validity of Koenker and Bassett's (1978) estimator. However, under the appropriate conditions, our estimator identifies the same conditional quantiles, the optimal predictors under the usual asymmetric loss function, and these are inherently robust.

The setup we consider is restrictive in that we need to assume that the covariates only affect the distribution of interest through known location and scale functions.[1] However, practitioners are often prepared to make even stronger assumptions,[2] and we will argue that in spite of its assumptions our approach can be useful in many empirical applications. Importantly, although we do not develop such tests here, it is possible to test the assumption that the covariates only affect the location and scale functions, and therefore it is possible to check whether or not our approach is suitable in a particular application.

The approach we propose is not meant to replace the well-established and very attractive estimation methods based on the check-function. Instead, we see our estimator as an additional tool that can complement those techniques and allow the estimation of regression quantiles in settings where otherwise that would be difficult or even impossible. For example, our approach is attractive when panel data are available and the researcher wants to estimate regression quantiles including individual effects.

Quantile regressions with individual effects suffer from the incidental parameters problem (see, e.g., Neyman and Scott, 1948, and Lancaster, 2000), and there is now a substantial literature dealing with the challenges posed by these models (see, e.g., Koenker, 2004, Lamarche, 2010, Canay, 2011, Galvão, 2011, Kato, Galvão and Montes-Rojas, 2012, Galvão and Wang, 2015, Galvão and Kato, 2016, and Powell, 2017). However, none of these methods gained widespread popularity, either because of their computational complexity or because they rely on very restrictive assumptions on how the fixed effects affect the quantiles. Albeit also based on a somewhat restrictive (but testable) assumption, our approach has the advantage of being very easy to implement even in very large problems and it allows the individual effects to affect the entire distribution,

[1]Notice that in a conditional location-scale model the regressors affect all higher-order moments through the scale function. Indeed, for $m > 1$, the m-th conditional central moment is proportional to the m-th power of the scale function.

[2]For example, the popular Tobit and sample selection models assume that the errors are normally distributed and statistically independent of the regressors.

rather than being just location shifters as in, e.g., Koenker (2004), Lamarche (2010), and Canay (2011).[3]

Our approach can also be adapted to the estimation of cross-sectional models with endogenous variables as, for example, in Abadie, Angrist, and Imbens (2002) and in Chernozhukov and Hansen (2005, 2006, and 2008). Strictly speaking, in this context our approach is not based on the estimation of conditional means, but on moment conditions that under exogeneity identify conditional means. The proposed estimator is closely related to that of Chernozhukov and Hansen (2008) in the sense that under suitable regularly conditions it identifies the same structural quantile function, but has the advantage of being applicable to non-linear models and being computationally much simpler, especially in models with multiple endogenous variables.

The remainder of the paper is organized as follows. Section 2 introduces our approach to the estimation of regression quantiles in location-scale models. Section 3 considers the application of our approach in the context of a panel data model with fixed effects. In Section 4 we consider estimation with cross-sectional data when some of the variables of the model are endogenous. Section 5 presents the results of a small simulation study and Section 6 illustrates the application of the proposed methods with two empirical examples. Section 7 concludes and an Appendix collects the more technical details.

2. THE BASIC IDEA

The rationale of the proposed estimator can be introduced in a simple setup. We are interested in estimating the conditional quantiles of a random variable Y whose distribution conditional on a k-vector of covariates X belongs to the location-scale family

[3] For example, in an application that motivated this work, our colleagues needed to estimate quantile regressions models with $14,000$ fixed effects and 70 other parameters using data on over $600,000$ individuals. Our estimator can easily deal with such large problems, but we are not aware of any other approach that can be used to estimate such models without restricting the fixed effects to be location shifters.

and can be expressed as

$$Y = \alpha + X'\beta + \sigma(\delta + Z'\gamma)U, \tag{1}$$

where:

- $(\alpha,\, \beta',\, \delta,\, \gamma')' \in \mathbb{R}^{2(k+1)}$ are unknown parameters;[4]
- Z is a k-vector of known differentiable (with probability 1) transformations of the components of X with element l given by

$$Z_l = \mathcal{Z}_l(X), \quad l = 1, \ldots, k;$$

- $\sigma(\cdot)$ is a known $\mathcal{C}^2$ function such that

$$P\{\sigma(\delta + Z'\gamma) > 0\} = 1;$$

- U is an unobserved random variable, independent of X, with density function $f_U(\cdot)$ bounded away from 0 and normalized to satisfy the moment conditions

$$E(U) = 0 \quad E(|U|) = 1. \tag{2}$$

A special case of (1) is, of course, the linear heteroskedasticity model in which $\sigma(\cdot)$ is the identity function and $Z = X$. This model has been study by many authors and has a long tradition in the quantile regression literature (see, e.g., Koenker and Basset, 1982, Gutenbrunner and Jurečková, 1992, Koenker and Zhao, 1994, He, 1997, and Zhao, 2000). Our formulation, however, is sufficiently general to also encompass other specifications such as models with multiplicative heteroskedasticity (Harvey, 1976), which have recently been advocated by Romano and Wolf (2017).

The specification in (1) differs from the standard formulation $Y = x'\beta(U)$, $U \sim \text{Uniform}(0,1)$, which can be viewed as representing a linear data generating process where all unobserved heterogeneity comes from random parameter variation and each

[4]For simplicity, we assume that X and Z have the same dimension.

parameter is allowed to be a different function of U. The model in (1) imposes that there is a single source of unobserved heterogeneity, but our formulation allows for nonlinear quantile effects and, thus, it cannot be considered a restricted version of $Y = x'\beta(U)$, except when $\sigma(\cdot)$ is the identity function. In this case (1) is also a linear model where all unobserved heterogeneity comes from random parameter variation, but the distributions of the coefficients are assumed to differ only in their location and scale.

Model (1) implies that

$$Q_Y(\tau|X) = \alpha + X'\beta + \sigma\left(\delta + Z'\gamma\right) q\left(\tau\right) \tag{3}$$

with $q\left(\tau\right) = F_U^{-1}(\tau)$, and therefore $\Pr\left(U < q\left(\tau\right)\right) = \tau$. In the case where $\sigma\left(\cdot\right)$ is the identity function and $Z = X$, the quantiles simplify to

$$Q_Y\left(\tau|X\right) = \left(\alpha + \delta q\left(\tau\right)\right) + X'\left(\beta + \gamma q\left(\tau\right)\right).$$

In general, the marginal effect of the regressor X_l on the τ-th quantile of Y (the "regression quantile coefficient") is

$$\beta_l(\tau, X) = \beta_l + q\left(\tau\right) D_{X_l}^{\sigma} \tag{4}$$

with $D_{X_l}^{\sigma} = \partial\sigma\left(\delta + Z'\gamma\right)/\partial X_l$.

Using (2), and the exogeneity of the regressors, the vector of parameters of interest, $(\alpha,\, \beta',\, \delta,\, \gamma',\, q(\tau))'$, can be identified from the following set of moment conditions (for ease of exposition we assume here *i.i.d.* data):

$$\begin{aligned} E[RX] &= 0 \\ E[R] &= 0 \\ E\left[\left(|R| - \sigma\left(\delta + Z'\gamma\right)\right) D_{\gamma}^{\sigma}\right] &= 0 \\ E\left[\left(|R| - \sigma\left(\delta + Z'\gamma\right)\right) D_{\delta}^{\sigma}\right] &= 0 \\ E[I(R \leq q\left(\tau\right)\sigma(\delta + Z'\gamma)) - \tau] &= 0 \end{aligned} \tag{MC1}$$

where $R = Y - (\alpha + X'\beta) = \sigma\left(\delta + Z'\gamma\right) U$, $D_{\gamma}^{\sigma} = \partial\sigma\left(\delta + Z'\gamma\right)/\partial\gamma$, $D_{\delta}^{\sigma} = \partial\sigma\left(\delta + Z'\sigma\right)/\partial\delta$.

Given that the location-scale model specifies the scale function $\sigma(\cdot)$, we can explore that information and base the identification on the alternative set of moment conditions

$$\begin{aligned} E[UX] &= 0 \\ E[U] &= 0 \\ E\left[(|U|-1)\, D^{\sigma}_{\gamma}\right] &= 0 \\ E\left[(|U|-1)\, D^{\sigma}_{\delta}\right] &= 0 \\ E\left[I\left(U < q\left(\tau\right)\right) - \tau\right] &= 0 \end{aligned} \tag{MC2}$$

where $U = \left(Y - \left(\alpha + X'\beta\right)\right)/\sigma\left(\delta + Z'\gamma\right)$.[5]

These conditions form the basis of the estimation procedure (Method of Moments-Quantile Regression, MM-QR) discussed in further detail in the next sections. Conditions (MC1) bear resemblance to those of the Restricted Quantile Regression of He (1997) and Zhao (2000) but we explore different moment conditions. In He (1997) and Zhao (2000) the moment conditions corresponding to (2) are that U has median at zero and that $|U|$ has median at 1. Thus, the implied orthogonality condition corresponding to (MC1) are those defining least absolute deviation estimators rather than least squares estimators. Our choice is, admittedly, weaker from a robustness point of view, but we believe that our approach is useful in that it makes it very easy to implement quantile regression in a wider class of models.[6] In particular, we will explore the use of (MC1) in the estimation of panel data models with fixed effects, and (MC2) in the estimation of structural quantile functions as defined by Chernozhukov and Hansen (2006, 2008).

[5] Although we do not pursue that here, it is easy to see that the validity of the location-scale model can be tested, for example, by testing the overidentifying restrictions resulting from augmenting (MC1) and (MC2) with conditions imposing the orthogonality between suitable functions of U and functions of the regressors. See, e.g., Hansen (1982) and Newey (1985).

[6] Notice that, due to the normalization in (2), we estimate the scale function rather than the skedastic function. There are two reasons for this. First, in the leading case where the scale is a linear function of the regressors and the quantiles are linear, the scale function can be estimated by ordinary least squares, whereas estimation of the skedastic function would involve non-linear estimation. Additionally, as noted by Koenker and Zhao (1996), the scale function is a more robust measure of dispersion.

3. PANEL DATA WITH FIXED EFFECTS

3.1. Linear models

The estimation of linear regression quantiles for longitudinal data was seminally considered by Koenker (2004). To mitigate the effects of the incidental parameters problem, Koenker considers a model where the individual effects only cause parallel (location) shifts of the distribution of the response variable (see also Lamarche, 2010, Canay, 2011, and Galvão, 2011). We also start by considering a linear specification, but allow the individual effects to affect the entire distribution, as in Kato, Galvão and Montes-Rojas (2012), Galvão and Wang (2015), and Galvão and Kato (2016).

Given data $\{(Y_{it},\, X_{it}')'\}$ from a panel of n individuals $i = 1, \ldots, n$ over T time periods, $t = 1, \ldots, T$, we consider the estimation of the conditional quantiles $Q_Y\left(\tau|X\right)$ for a location-scale model of the form

$$Y_{it} = \alpha_i + X_{it}'\beta + (\delta_i + Z_{it}'\gamma)U_{it}, \tag{5}$$

with $P\{\delta_i + Z_{it}'\gamma > 0\} = 1$. The parameters (α_i, δ_i), $i = 1, \ldots, n$, capture the individual i fixed effects and Z is defined as before. The sequence $\{X_{it}\}$ is *i.i.d.* for any fixed i and independent across t. U_{it} are *i.i.d.* (across i and t), statistically independent of X_{it}, and normalized to satisfy the moment conditions (2).[7]

Model (5) implies that

$$Q_Y\left(\tau|X_{it}\right) = \left(\alpha_i + \delta_i q\left(\tau\right)\right) + X_{it}'\beta + Z_{it}'\gamma q\left(\tau\right). \tag{6}$$

We will call the scalar coefficient $\alpha_i(\tau) \equiv \alpha_i + \delta_i q\left(\tau\right)$ the quantile-τ fixed effect for individual i, or the distributional effect at τ. The distributional effect differs from the usual fixed effect in that it is not, in general, a location shift. That is, the distributional effect represents the effect of time-invariant individual characteristics which, like other variables, are allowed to have different impacts on different regions of the conditional

[7] Notice that these conditions do not imply strict exogeneity.

distribution of Y. The fact that $\int_0^1 q(\tau)\, d\tau = 0$ implies that α_i can be interpreted as the average effect for individual i.

Consider now the MM-QR estimator of (6) implied by (MC1). For this model, the moment conditions have a convenient triangular structure with respect to the model parameters that allows the one-step GMM estimator (Hansen, 1982) to be calculated sequentially:

1. Regress $(Y_{it} - \sum_t Y_{it}/T)$ on $(X_{it} - \sum_t X_{it}/T)$ by least squares to obtain $\hat{\beta}$;

2. Estimate $\hat{\alpha}_i = \frac{1}{T}\sum_t (Y_{it} - X_{it}'\hat{\beta})$ and obtain the residuals $\hat{R}_{it} = Y_{it} - \hat{\alpha}_i - X_{it}'\hat{\beta}$;

3. Regress $(|\hat{R}_{it}| - \sum_t |\hat{R}_{it}|/T)$ on $(Z_{it} - \sum_t Z_{it}/T)$ by least squares to obtain $\hat{\gamma}$;

4. Estimate $\hat{\delta}_i = \frac{1}{T}\sum_t (|\hat{R}_{it}| - Z_{it}'\hat{\gamma})$;

5. Estimate $q(\tau)$ by $\hat{q}$, the solution to

$$\min_q \sum_i \sum_t \rho_\tau \left(\hat{R}_{it} - \left(\hat{\delta}_i + Z_{it}'\hat{\gamma} \right) q \right)$$

   where $\rho_\tau(A) = (\tau - 1)AI\{A \leq 0\} + \tau AI\{A > 0\}$ is the check-function. (Equivalently, order the standardized residuals $\hat{U} = \hat{R}_{it} / \left(\hat{\delta}_i + Z_{it}'\hat{\gamma} \right)$ and estimate the τ-th sample quantile.)

The regression in Step 3 is reminiscent of the one used to compute Glejser's (1969) test for heteroskedasticity, and the insights in Machado and Santos Silva (2000) and Im (2000) suggest that the MM-QR estimator is greatly simplified if $|R|$ in (MC1) is replaced by

$$2R(I\{R \geq 0\} - P\{R \geq 0\}).$$

Indeed, because $|R| = 2R(I\{R \geq 0\} - 1/2)$, the two transformations differ only in the way the residuals R are weighted: with mean zero in one case and with mean $P\{R \geq 0\} - 1/2$ in the other. Using the assumption that $E[R|Z] = 0$, it is clear that the (population) moment condition

$$E[Z\,(|R| - \delta_i - Z'\gamma)] = 0$$

identifies δ_i and γ *iff*

$$E[Z\,(2R(I\{R \geq 0\} - \eta) - \delta_i - Z'\gamma)] = 0, \quad \eta = P(R \geq 0) = P(U \geq 0).$$

Therefore, in Steps 3 and 4, instead of using $|\hat{R}|$ one may use

$$2\hat{R}_{it}[I\{\hat{R}_{it} \geq 0\} - \hat{\eta}]$$

with

$$\hat{\eta} = \frac{1}{nT}\sum_i \sum_t I\{\hat{R}_{it} \geq 0\}.$$

The advantage of using this alternative transformation in Steps 3 and 4 of the algorithm is that it makes asymptotic inference on γ independent of the first step estimator. Besides simplifying the treatment of the asymptotic properties of the estimator, this allows the practitioner to make inference about the parameters of the scale function directly from the least squares results in the modified Step 3, without having to take into account the fist-step estimation.

Below we present the main results on the asymptotic properties of the MM-QR estimator as a set of theorems whose proofs are provided in the Appendix. The following results could be obtained using a standard GMM framework for the exactly identified case and the results of, say, Newey and McFadden, (1994, Theorem 7.2). Our approach however, mimics the sequence of steps above and is similar to Zhao's (2000). Throughout we use the following notation: for any sequence of random variables A_{it}, B_{it} for which the limits exist,

$$Q_{AB} = \lim_{n\to\infty} \frac{1}{n} \sum_i E[(A_{i1} - \mu_{A_i})(B_{i1} - \mu_{B_i})']$$

with $\mu_{A_i} = E[A_{it}]$,

$$P_{AB} = \lim_{n\to\infty} \frac{1}{n} \sum_i E[\sigma_{i1}^2 (A_{i1} - \mu_{A_i})(B_{i1} - \mu_{B_i})']$$

with $\sigma_{it} = (\delta_i + Z_{it}'\gamma)$, and

$$P_A = \lim_{n\to\infty} \frac{1}{n} \sum_i E[\sigma_{i1}^2 (A_{i1} - \mu_{A_i})].$$

Our first theorem establishes the consistency of the MM-QR estimators.

Theorem 1 (Consistency) *Consider model (5) satisfying conditions (P) in the Appendix. Assume further that the sequences* $\{X_{it}, Z_{it}, U_{it}\}$ *satisfy the conditions (U) and (XZ) in the Appendix. Then, as* $(n, T) \to \infty$

$$\hat{\beta} - \beta \xrightarrow{P} 0$$
$$\hat{\gamma} - \gamma \xrightarrow{P} 0$$
$$\hat{q}(\tau) - q(\tau) \xrightarrow{P} 0.$$

□

It is easy to establish that the estimators of the intercepts α_i and δ_i are also consistent provided that $T \to \infty$. Furthermore, Lemmata 1 and 4 in the Appendix prove that if $n/T \to 0$ as $(n, T) \to \infty$, then

$$\max_{1 \leq i \leq n} |\hat{\alpha}_i - \alpha_i| = o_P(1)$$

and

$$\max_{1 \leq i \leq n} |\hat{\delta}_i - \delta_i| = o_P(1).$$

Next we establish the asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$.

Theorem 2 (Asymptotic distribution) *Consider model (5) satisfying conditions (P) in the Appendix. Assume further that the sequences* $\{X_{it}, Z_{it}, U_{it}\}$ *satisfy the conditions (U) and (XZ) in the Appendix. Then, as* $(n, T) \to \infty$

$$\sqrt{nT}(\hat{\beta} - \beta) \xrightarrow{\mathcal{D}} Q_{XX}^{-1} \mathcal{N}(0, E(U^2) P_{XX})$$

and if $(n, T) \to \infty$ *with* $n = o(T)$,

$$\sqrt{nT}(\hat{\gamma} - \gamma) \xrightarrow{\mathcal{D}} Q_{ZZ}^{-1} \mathcal{N}(0, E(V^2) P_{ZZ})$$

with $V = 2U(I\{U \geq 0\} - P\{U \geq 0\})$. □

Notice that, as is well known, the results for the least squares estimator $\hat{\beta}$ also hold when $n \to \infty$ for fixed T, or $T \to \infty$ for fixed n.[8] Also, as mentioned before, the limiting distribution of $\hat{\gamma}$ does not depend on the first-step estimation.

[8] This would not be the case if the estimator was based on (MC2).

It is not difficult to establish that the MM-QR estimators of the fixed effects coefficients α_i, δ_i converge at rate root-T to a Gaussian distribution. Owing to the faster rates of convergence of the slope estimators, this asymptotic distribution of the fixed effects coefficients is the same as if the slopes were known.

The quantile-τ fixed effect, $\alpha_i(\tau) = \alpha_i + \delta_i q(\tau)$, can be estimated by

$$\hat{\alpha}_i(\tau) = \frac{1}{T}\sum_{t=1}^{T}(Y_{it} - X_{it}'\hat{\beta}) + \hat{q}\frac{1}{T}\sum_{t=1}^{T}(|\hat{R}_{it}| - Z_{it}'\hat{\gamma}).$$

Likewise, the τ-th quantile regression coefficient of the regressor X_l, which is given by (4) and is the main parameter of interest, can be estimated by $\hat{\beta}_l(\tau, X) = \hat{\beta}_l + \hat{q}\hat{\gamma}$.

The consistency of $\hat{\beta}_l(\tau, X)$ follows directly from Theorem 1, and Theorem 3 establishes the asymptotic distribution of $\hat{\beta}_l(\tau, X)$ for the leading case where $Z = X$ and $\hat{\beta}_l(\tau, X) = \hat{\beta}_l(\tau)$; the more general case is equally straightforward.[9]

Theorem 3 (Asymptotic distribution of the QR coefficients) *Consider model (5) satisfying conditions (P) in the Appendix and assume that $Z = X$. Assume further that the sequences $\{X_{it}, Z_{it}, U_{it}\}$ satisfy the conditions in (U) and (XZ) in the Appendix. Then, as $(n, T) \to \infty$ with $n = o(T)$*

$$\sqrt{nT}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{\mathcal{D}} \Xi \mathcal{N}(0, \Omega)$$

with

$$\Xi = \begin{pmatrix} Q_{XX}^{-1} & q(\tau)\, Q_{ZZ}^{-1} & (1/\mu_\sigma)\gamma \end{pmatrix},$$

being a $k \times (2k+1)$ matrix with blocks Q_{XX}^{-1}, $q(\tau)\, Q_{ZZ}^{-1}$, and $(1/\mu_\sigma)\gamma$, and

$$\Omega = \begin{pmatrix} E[U^2]P_{XX} & E[UV]P_{XZ} & E[UW]P_X \\ & E[V^2]P_{ZZ} & E[VW]P_Z \\ & & \mu_{\sigma^2}E[W^2] \end{pmatrix},$$

with $\mu_{\sigma^a} = \frac{1}{n}\sum_i(\delta_i + \gamma E[Z_{i1}])^a$, $a = 1, 2$ and $W = \frac{1}{f_U(q(\tau))}\psi_\tau(U - q(\tau)) - U - q(\tau)\, V$, where $\psi_\tau(A) = (\tau - I\{A \leq 0\})$. □

[9] With $Z = X$, $Q_{ZZ} = Q_{XX}$, and $P_{ZZ} = P_{ZX} = P_{XX}$; if $Z \neq X$, Ξ has to be adjusted in a straightforward way.

As with other quantile regression estimators for models with fixed effects (see, Galvão and Kato, 2018, and the references therein), the asymptotic distribution of $\hat{\beta}(\tau)$ has mean zero only when $(n, T) \to \infty$ with $n = o(T)$. If these conditions do not hold, the asymptotic distribution will be biased because, for fixed T, the variance of the estimator vanishes with n but the bias does not. Hence, as noted by Hahn and Newey (2004), confidence intervals may have poor coverage in applications where n/T is large. Because $\hat{\beta}$ is consistent even when T is fixed, the bias in the asymptotic distribution of $\hat{\beta}(\tau)$ comes from the fixed-T biases of $\hat{\gamma}$ and $\hat{q}(\tau)$. The next result sheds light on the nature of these biases by calculating the probability limits of $\hat{\gamma}$ and $\hat{q}(\tau)$ as n grows with T fixed.

Theorem 4 (Fixed T asymptotic biases) *Consider model (5) satisfying conditions (P) in the Appendix. Assume that $Z = X$ and that the sequences $\{X_{it}, Z_{it}, U_{it}\}$ satisfy the conditions in (U) and (XZ) in the Appendix. Assume further that the conditions in Lemma 5 in the Appendix are satisfied. Then, as $n \to \infty$ with T fixed*

$$\hat{\gamma} \xrightarrow{P} \gamma_T$$

with

$$\gamma_T = \gamma + \frac{1}{T} B^{\gamma}_{nT} + O(1/T^2),$$

and

$$\hat{q}(\tau) \xrightarrow{P} q_T(\tau)$$

with

$$q_T(\tau) = q(\tau) + \frac{1}{T} B^{q}_{nT} + O(1/T^2),$$

where both B^{γ}_{nT} and B^{q}_{nT} are $O_P(1)$ as $n \to \infty$ and $B^{\gamma}_{nT} = 0$ when $\gamma = 0$. □

This result shows that it is possible to use jackknife bias corrections to eliminate the $O\left(T^{-1}\right)$ term in the biases of $\hat{\gamma}$ and $\hat{q}\left(\tau\right)$ (see Hahn and Newey, 2004, Dhaene and Jochmans, 2015, and Fernández-Val and Weidner, 2016). The bias-corrected estimates of γ and $q(\tau)$ can then be used to eliminate the leading term in the bias of $\hat{\beta}(\tau)$, thereby substantially attenuating the inference problems identified by Hahn and Newey (2004). In Section 5 we present simulation results for a range of values of n, T, and n/T and find that, as expected, the confidence intervals have poor coverage when n/T is large (above 10). However, our results also show that the coverage of the confidence intervals is much improved by using the simple split-panel jackknife bias correction of Dhaene and Jochmans (2015).

In summary, the proposed estimator suffers from the incidental parameters problem and, in that sense, it has no advantage over alternative approaches. However, because we can partial out the fixed effects in the estimation of β and γ, our estimator is computationally much easier to implement than any other estimator for quantile regression models with fixed effects. Indeed, our estimator is as easy to implement as the popular "within" estimator and remains practical even for models with many regressors estimated with samples where n is very large. Moreover, with the easy to implement jackknife bias correction, it allows reasonably reliable inference to be performed for moderate values of T, even when n/T is large.

We conclude this sub-section by noting that the estimates of the conditional quantiles obtained from (MC1) do not cross (see also He, 1997); this follows directly from the unidimensional nature of the quantile estimator implied by the last moment condition of (MC1). The following proposition formally establishes this result for the estimator consider here, and similar results can be straightforwardly obtained for the quantiles of other location-scale models evaluated at estimates obtained from (MC1) or (MC2).

Proposition 1 (No Quantile-Crossing: He, 1997) *Let $Z = X$ and consider the regression quantile $Q_Y(\tau|X)$ given by (6) and its estimate $\hat{Q}_Y(\tau|X) = \hat{\alpha}_i + X_{it}'\hat{\beta} + \left(\hat{\delta}_i + X_{it}'\hat{\gamma}\right)\hat{q}(\tau)$. Then, for any design point with $(\hat{\delta}_i + X_{it}'\hat{\gamma}) > 0$,*

$$\tau \leq \tau' \Leftrightarrow \hat{Q}_Y(\tau|X) \leq \hat{Q}_Y(\tau'|X).$$

□

3.2. Non-linear models

The linear heteroskedasticity model considered so far is particularly attractive for its long history and for its simplicity, but estimation with other specifications of the location and scale functions is also possible. However, in specifications with fixed effects, estimating non-linear models will generally be impractical.

The exception to this are specifications based on the exponential function because in this case, just like in the linear model, there is a transformation that eliminates the fixed effects. Indeed, Wooldridge (1999) shows that the so-called fixed effects Poisson regression with an exponential conditional mean, which conditions-out the individual effects, is valid under very general conditions and is easy to implement (notice that this estimator is valid even if the data are not counts).

The possibility of estimating models with $\sigma(\cdot) = \exp(\cdot)$ is particularly interesting because this specification ensures that $\sigma(\cdot) > 0$. Moreover, models with multiplicative heteroskedasticity also have a long history and are popular in many contexts (see, e.g., Harvey, 1976, Wooldridge, 2010, and Romano and Wolf, 2017).[10]

Therefore, when either the conditional mean, the conditional variance, or both, are given by exponential functions, all that is needed is to replace the corresponding least squares steps in the algorithm described before with suitable Poisson regressions; naturally, the subsequent computation of the fixed effects needs to be modified accordingly, but that is trivial.

Using the delta-method and our earlier results, it is possible to derive the asymptotic distribution of the estimators in these non-linear models. Notice, however, that in non-linear models the regression quantile coefficients will depend on the estimates of the fixed effects; for example, in a linear model with multiplicative heteroskedasticity, the regression quantile coefficients for individual i will depend on $\hat{\delta}_i$. In practice, we can

[10] For example, the latest release of Stata (StataCorp., 2017) includes the command `hetregress` which estimates linear regression models with multiplicative heteroskedasticity.

either take this into account when applying the delta method, or we may obtain results conditioning on a given value of the fixed effect, such as the sample average of $\hat{\delta}_i$.

4. ENDOGENOUS REGRESSORS

We explore now the application of the MM-QR estimator to cross-sectional models with endogenous explanatory variables. Consider a scalar random variable Y that is related to an unobserved scalar random variable U satisfying (2) and to a vector of observed random variables $(D', C_1', C_2')'$ (with dimensions k_D, k_1, k_2, respectively, and $k_2 \geq k_D$), by the following structural relationship

$$
\begin{aligned}
Y &= D'\beta_D + C_1'\beta_1 + \sigma\left(D'\gamma_D + C_1'\gamma_1\right)U \\
D_l &= \mathcal{D}_l\left(C_1, C_2, U^{\star}\right) \text{ for } l = 1, \ldots, k_D \\
&C_1, C_2 \text{ statistically independent of } U,
\end{aligned}
\tag{7}
$$

where $\mathcal{D}_l(\cdot) : \mathbb{R}^{k_1+k_2+1} \to \mathbb{R}$, $\sigma(\cdot)$ is as defined in Section 2, and $U^{\star}$ is an unobserved random variable that may not be independent of U. The parameters $(\beta, \gamma) \in \Omega_2$, satisfy assumption (P1) in the Appendix. Put $X' = (D', C_1')$ (the regressors), $C' = (C_1', C_2')$ (the instruments), $\beta' = (\beta_D', \beta_1')$, and $\gamma' = (\gamma_D', \gamma_1')$.[11]

The most relevant feature of this model is that the endogenous regressor impacts both the location and scale of Y. Although similar, (7) is neither more nor less restrictive than the structural random coefficients model considered by Chernozhukov and Hansen (2006, 2008). As noted before, in the linear case we impose that, up to location and scale, all coefficients have the same distribution, whereas Chernozhukov and Hansen (2006, 2008) allow each coefficient to have different distributions. However, unlike them, we allow for non-linear quantile effects.

As in Chernozhukov and Hansen (2006, 2008), we are not interested in estimating $Q_Y(\tau|\,X)$, but the parameters of a function $S_Y(\tau|\,X)$ such that

$$P\{Y \leq S_Y(\tau|\,X)\} = P\{Y \leq S_Y(\tau|\,X)|\,C\} = \tau.$$

[11] Notice that if the location and scale have intercepts, C_1 will have a column of 1s.

Therefore, $S_Y(\tau|\, X)$, the "structural quantile function" in Chernozhukov and Hansen's (2008) terminology, can be interpreted as $Q_Y\left(\tau|C\right)$ and can be written as

$$S_Y(\tau|\, X) = X'\beta + \sigma(X'\gamma)q(\tau).$$

Given the model, if (β,γ) were known, the moment condition

$$E\left[\psi_\tau\left(\frac{Y - X'\beta}{\sigma(X'\gamma)} - q\right)\right] = 0$$

would identify the marginal quantile of U, that is $q(\tau)$ such that $P\{U \leq q(\tau)\} = P\{U \leq q(\tau)|\, C\} = \tau$. This procedure is not feasible since β and γ are not known but, given the data $\{(Y_i,\, X_i', C_i')'\}$, these parameters can be consistently estimated under very general conditions by applying GMM to the sample analogues of the moment conditions in (MC2),

$$\frac{1}{\sqrt{n}}\sum_1^n C_i\left(\frac{Y_i - X_i'\hat{\beta}}{\sigma(X_i'\hat{\gamma})}\right) = 0,$$
$$\frac{1}{\sqrt{n}}\sum_1^n C_i\left(\frac{\left|Y_i - X_i'\hat{\beta}\right|}{\sigma(X_i'\hat{\gamma})} - 1\right) = 0.$$

Notice that this MM-QR estimator cannot be solved sequentially, and therefore in this case there is no practical benefit in replacing $|U_i|$ with $2U_i\left(I\left\{U_i > 0\right\} - P\left\{U > 0\right\}\right)$.

Given the estimates of β and γ, $q(\tau)$ may be estimated by the condition

$$\frac{1}{\sqrt{n}}\sum_1^n \psi_\tau\left(\frac{Y_i - X_i'\hat{\beta}}{\sigma(X_i'\hat{\gamma})} - q\right) = o_P(1)$$

or, alternatively, by ranking the standardized residuals.

The next theorem formalizes this estimator for the exactly identified case ($k_D = k_2$); the over-identified case could be handled similarly.[12]

Theorem 5 (Structural quantile function coefficients) *Consider a sample of* n *i.i.d.. observations of* (Y, X, C) *from the structure defined by (7) with* $\dim\left(X\right) = \dim\left(C\right)$.

[12] In the Appendix, we present a generalization of this result for the case where multiple quantiles are estimated. A similar generalization of the results in Theorem 3 is also straightforward.

Then, under assumptions (P), (U), and (DC) in the Appendix, as $n \to \infty$

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}-\beta) \\ \sqrt{n}(\hat{\gamma}-\gamma) \\ \sqrt{n}(\hat{q}-q(\tau)) \end{pmatrix} \xrightarrow{\mathcal{D}} G^{-1}\mathcal{N}(0,\Omega),$$

where,

$$\Omega = \begin{pmatrix} E[U^2]E[CC'] & E[UV]E[CC'] & \frac{E[U\psi_\tau(U-q(\tau))}{f_U(q(\tau))}E[C] \\ & E[V^2]E[CC'] & \frac{E[V\psi_\tau(U-q(\tau))}{f_U(q(\tau))}E[C] \\ & & \frac{1}{f_U^2(q(\tau))}\tau(1-\tau) \end{pmatrix},$$

with $V = |U| - 1$ *and*

$$G = \begin{pmatrix} E[(1/\sigma)\,CX'] & E[(\sigma'/\sigma)\,UCX'] & 0_{k\times 1} \\ E[(1/\sigma)\,\mathrm{sign}(U)CX'] & E[(\sigma'/\sigma)\,|U|CX'] & 0_{k\times 1} \\ E[(1/\sigma)\,X'] & E[(\sigma'/\sigma)\,UX'] & 1 \end{pmatrix},$$

with $k = k_1 + k_2$*,* $\sigma = \sigma(X'\gamma)$*, and* $\sigma' = d\sigma(z)/dz$ *at* $z = X'\gamma$*.* □

Inference about $\beta(\tau, X) = \partial S_Y(\tau|\, X)/\partial X$, the ultimate parameter of interest, can be performed using the standard delta-method. For example, in the linear case where $\beta(\tau, X) = \beta(\tau) = \beta + \gamma q\,(\tau)$ we have that

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{\mathcal{D}} AG^{-1}\mathcal{N}(0,\Omega),$$

where

$$A = \begin{pmatrix} \mathcal{I}_{k\times k} & q\,(\tau)\,\mathcal{I}_{k\times k} & \gamma \end{pmatrix}$$

is a $k \times (2k+1)$ matrix with blocks $\mathcal{I}_{k\times k}$, $q\,(\tau)\,\mathcal{I}_{k\times k}$, and γ, where $\mathcal{I}_{k\times k}$ denotes a $k \times k$ identity matrix.

Our approach to the estimation of the structural quantile function can be seen as a contribution to the growing literature addressing the computational challenges faced in the implementation of the Chernozhukov and Hansen (2008) estimator. Although several promising approaches to this problem have been developed, as far as we know, all of them have unappealing features such as requiring the tuning of the optimization algorithm (Chernozhukov and Hong, 2003), the selection of tolerance parameters (Xu and Burer, 2017), the choice of a smoothing parameter (Kaplan and Sun, 2017), the specification of the parameter space (Chen and Lee, 2017), or being limited to models with binary treatments (Wüthrich, 2015). In contrast, our estimator is extremely simple to implement, even if the model is non-linear and has multiple endogenous explanatory variables, and it ensures that the estimated structural quantile functions do not cross.

Therefore, at the very least, the proposed estimator can be useful to provide starting values for other methods and to guide in the definition of the parameter space. In the next sections we present simulation results and an empirical example illustrating the performance and application of this estimator.

5. SIMULATION EVIDENCE

This section presents the results of two small simulation exercises illustrating the performance of the methods proposed in Sections 3 and 4.

5.1. Panel data models with fixed effects

The first set of experiments is designed to study the performance of the estimator in a panel-data model with fixed effects. For this experiment, 10,000 independent data sets were generated as

$$Y_{it} = \alpha_i + X_{it} + (1 + X_{it} + \kappa\alpha_i)\, U_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{8}$$

where $\alpha_i \sim \chi^2_{(1)}$ and $X_{it} = 0.5\,(\alpha_i + \chi_{it})$, with $\chi_{it} \sim \chi^2_{(1)}$, and three different distributions of U_{it} are considered: $\mathcal{N}(0,1)$, $\chi^2_{(5)}$, and $t_{(5)}$; in all cases U_{it} is standardized to have zero mean and unit variance.[13] We performed simulations for $T \in \{10, 20, 50\}$, $n \in \{50, 500, 100T\}$, $\tau \in \{0.25, 0.75\}$, and $\kappa \in \{0, 1\}$. For $\kappa = 0$ the fixed effects are pure location shifts as assumed by Koenker (2004) and Canay (2011); otherwise the fixed effects affect the entire distribution.

[13] Using this normalization rather than $E\,|U_{it}| = 1$ is immaterial and facilitates the data generation.

The MM-QR estimator described in Section 3 was used to estimate linear quantile regressions for these data and Tables 1 and 2 report the bias, standard error (SE), and mean squared error (MSE) for all the cases with $\tau = 0.25$; we do not report the results obtained with $\tau = 0.75$ because they lead to similar conclusions. The tables also report the results obtained with the bias-corrected version of the estimator based on the split-panel jackknife of Dhaene and Jochmans (2015), these results are labeled JKBC.[14]

The results in Tables 1 and 2 confirm that the bias of the MM-QR estimator drops as T grows, being essentially proportional to $1/T$. A notable feature of the results in Tables 1 and 2 is that the jackknife bias correction is extremely effective.[15] Indeed, even for the smallest values of n and T considered, the jackknife correction essentially eliminates the bias without a significant loss of precision.

Our results also confirm that the precision of the estimators increases with nT and this is reflected in the values of SE and MSE. As noted before, the fact that the bias decreases with T while the variance decreases with nT may lead the asymptotic distribution of the estimator to be biased when n/T is large (see Hahn and Newey, 2004).

To investigate the extent of this problem, we used an estimator of the covariance matrix presented in Theorem 3 to compute 95% confidence intervals centred at the MM-QR estimates and at their bias-corrected counterparts; Table 3 displays the coverage rates of these intervals. These results suggest that for n/T up to 10 the coverage of the confidence intervals centered at the MM-QR estimates is reasonable,[16] but for larger

[14] We also estimated the model using Canay's (2011) estimator; for brevity we do not present these results in detail but briefly discuss them now. When $\kappa = 0$, Canay's estimator imposes the valid restriction that the fixed effects are pure location shifters and consequently has lower SE than the MM-QR estimator and often has somewhat smaller bias; the performance of the two estimators is, however, comparable even for non-normal data. Naturally, the performance of Canay's estimator deteriorates sharply when $\kappa = 1$, which reflects the sensitivity of the estimator to departures from its key assumption.

[15] Because the estimator of β is unbiased, we implement the estimator by adding to $\hat{\beta}$ the product of bias-corrected estimates of γ and q.

[16] Following Cochran (1952), we consider departures from the nominal 95% coverage to be unimportant if the estimated coverage is between 0.9354 and 0.9638.

values of n/T the coverage rates can drop dramatically; this is especially clear in Case 2. However, centering the intervals at the bias-corrected estimates greatly alleviates the problem, generally leading to intervals with good coverage. We note, however, that in some cases these intervals can be too wide, with coverage of about 99%.

Table 1: Bias, SE, and MSE results for $\tau = 0.25$ and $\kappa = 0$

		$n = 50$		$n = 500$		$n = 100 \times T$	
T		MMQR	JKBC	MMQR	JKBC	MMQR	JKBC
				Case 1: $\mathcal{N}(0,1)$			
10	BIAS	0.099	0.008	0.079	−0.006	0.079	−0.006
	SE	0.316	0.343	0.103	0.110	0.073	0.077
	MSE	0.109	0.118	0.017	0.012	0.011	0.006
20	BIAS	0.047	0.000	0.038	−0.002	0.036	−0.002
	SE	0.231	0.243	0.073	0.076	0.037	0.038
	MSE	0.056	0.059	0.007	0.006	0.003	0.001
50	BIAS	0.016	−0.001	0.014	−0.000	0.014	−0.000
	SE	0.148	0.151	0.047	0.047	0.015	0.015
	MSE	0.022	0.023	0.002	0.002	0.000	0.000
				Case 2: $\chi^2_{(5)}$			
10	BIAS	0.149	0.013	0.131	0.003	0.130	0.002
	SE	0.225	0.241	0.071	0.075	0.050	0.053
	MSE	0.073	0.058	0.022	0.006	0.019	0.003
20	BIAS	0.075	0.002	0.066	0.001	0.065	−0.000
	SE	0.160	0.168	0.049	0.051	0.025	0.026
	MSE	0.031	0.028	0.007	0.003	0.005	0.001
50	BIAS	0.031	0.001	0.027	0.000	0.026	0.000
	SE	0.099	0.102	0.031	0.032	0.010	0.010
	MSE	0.011	0.010	0.002	0.001	0.001	0.000
				Case 3: $t_{(5)}$			
10	BIAS	0.062	−0.000	0.047	−0.010	0.046	−0.010
	SE	0.333	0.362	0.105	0.112	0.073	0.078
	MSE	0.115	0.131	0.013	0.013	0.008	0.006
20	BIAS	0.026	−0.003	0.020	−0.004	0.019	−0.004
	SE	0.228	0.236	0.074	0.076	0.037	0.038
	MSE	0.052	0.056	0.006	0.006	0.002	0.001
50	BIAS	0.009	−0.000	0.008	0.000	0.007	−0.001
	SE	0.147	0.150	0.047	0.047	0.015	0.015
	MSE	0.022	0.022	0.002	0.002	0.000	0.000

Table 2: Bias, SE, and MSE results for $\tau = 0.25$ and $\kappa = 1$

		$n = 50$		$n = 500$		$n = 100 \times T$	
T		MMQR	JKBC	MMQR	JKBC	MMQR	JKBC
		Case 1: $\mathcal{N}(0,1)$					
10	BIAS	0.101	0.008	0.080	−0.005	0.080	−0.004
	SE	0.416	0.452	0.134	0.144	0.094	0.101
	MSE	0.183	0.205	0.024	0.021	0.015	0.010
20	BIAS	0.048	−0.000	0.038	−0.002	0.037	−0.002
	SE	0.301	0.315	0.095	0.098	0.048	0.049
	MSE	0.093	0.099	0.010	0.010	0.004	0.002
50	BIAS	0.016	−0.002	0.015	−0.000	0.015	0.000
	SE	0.190	0.194	0.060	0.061	0.019	0.019
	MSE	0.036	0.038	0.004	0.004	0.001	0.000
		Case 2: $\chi^2_{(5)}$					
10	BIAS	0.149	0.010	0.129	0.000	0.128	0.000
	SE	0.296	0.321	0.093	0.100	0.065	0.070
	MSE	0.110	0.103	0.025	0.010	0.021	0.005
20	BIAS	0.074	0.001	0.065	−0.000	0.063	−0.001
	SE	0.208	0.220	0.064	0.067	0.032	0.034
	MSE	0.049	0.048	0.008	0.005	0.005	0.001
50	BIAS	0.030	0.001	0.026	−0.000	0.026	0.000
	SE	0.128	0.131	0.040	0.041	0.013	0.013
	MSE	0.017	0.017	0.002	0.002	0.001	0.000
		Case 3: $t_{(5)}$					
10	BIAS	0.064	0.001	0.049	−0.008	0.048	−0.007
	SE	0.437	0.473	0.136	0.145	0.094	0.101
	MSE	0.195	0.223	0.021	0.021	0.011	0.010
20	BIAS	0.026	−0.003	0.021	−0.004	0.021	−0.003
	SE	0.295	0.306	0.095	0.097	0.048	0.049
	MSE	0.088	0.093	0.009	0.010	0.003	0.002
50	BIAS	0.009	−0.001	0.009	0.000	0.008	−0.000
	SE	0.190	0.193	0.061	0.061	0.019	0.019
	MSE	0.036	0.037	0.004	0.004	0.000	0.000

Overall these simulation results are encouraging in that they suggest that the MM-QR estimator of the quantile regression model with fixed effects may be reasonably well behaved in many empirical applications, especially when its bias-corrected version is used.

Table 3: Coverage rates of 95% confidence intervals with $\tau = 0.25$

		$n = 50$		$n = 500$		$n = 100 \times T$	
	T	MMQR	JKBC	MMQR	JKBC	MMQR	JKBC
				Case 1: $\mathcal{N}(0,1)$			
$\kappa = 0$	10	0.9445	0.9473	0.9180	0.9615	0.8757	0.9680
	20	0.9563	0.9566	0.9413	0.9624	0.8725	0.9595
	50	0.9633	0.9620	0.9610	0.9675	0.8835	0.9665
				Case 2: $\chi^2_{(5)}$			
	10	0.9505	0.9727	0.7984	0.9842	0.6192	0.9902
	20	0.9661	0.9782	0.8825	0.9861	0.4783	0.9850
	50	0.9818	0.9856	0.9541	0.9876	0.4512	0.9866
				Case 3: $t_{(5)}$			
	10	0.9454	0.9443	0.9400	0.9583	0.9309	0.9628
	20	0.9545	0.9539	0.9495	0.9522	0.9308	0.9535
	50	0.9610	0.9584	0.9558	0.9566	0.9359	0.9586
				Case 1: $\mathcal{N}(0,1)$			
$\kappa = 1$	10	0.9543	0.9530	0.9360	0.9620	0.9125	0.9672
	20	0.9594	0.9583	0.9541	0.9654	0.9123	0.9619
	50	0.9650	0.9648	0.9649	0.9678	0.9176	0.9684
				Case 2: $\chi^2_{(5)}$			
	10	0.9674	0.9748	0.8895	0.9836	0.7826	0.9881
	20	0.9776	0.9819	0.9403	0.9864	0.7199	0.9862
	50	0.9860	0.9869	0.9723	0.9886	0.7108	0.9885
				Case 3: $t_{(5)}$			
	10	0.9527	0.9469	0.9488	0.9585	0.9441	0.9600
	20	0.9610	0.9599	0.9565	0.9554	0.9422	0.9564
	50	0.9635	0.9606	0.9584	0.9588	0.9453	0.9601

5.2. Cross-sectional model with endogeneity

The second set of experiments was designed to study the behavior of the MM-QR estimator for a cross-sectional model with an endogenous explanatory variable. In this case, $10,000$ independent cross-sectional data sets were simulated from

$$Y_i = 1 + D_i + (1 + D_i)\, U_i, \quad i = 1, \ldots, N, \tag{9}$$

with $D_i = ((1-\lambda)\, C_i + \lambda\, |U_i|)$, where $0 < \lambda < 1$ is a parameter, $C_i = |\xi_i|$, ξ_i has the same distribution as U_i and is independent of it, and again we consider three different distributions for the error: $\mathcal{N}(0,1)$, $\chi^2_{(5)}$, and $t_{(5)}$; in all cases U_i is standardized to have zero mean and unit variance. In this design D_i is endogenous and C_i is a valid instrument for it. Because of the endogeneity, the distribution of D_i necessarily varies with the distribution of U_i; we also let the distribution of C_i vary with the distribution of U_i so that the strength of the instrument depends only on the parameter λ. We performed simulations for $n \in \{200, 1000, 5000\}$, $\tau \in \{0.25, 0.75\}$, and $\lambda \in \{0.50, 0.25\}$.

We estimate structural quantile functions for (9) using the MM-QR estimator described in Section 4 and, for comparison, we also estimate the models using the IVQR estimator of Chernozhukov and Hansen (2008).[17] Table 4 reports the bias, standard error (SE), and mean squared error (MSE) for all the cases in this set of experiments for which $\tau = 0.25$; as before, we do not report the results with $\tau = 0.75$ which lead essentially to the same conclusions.

Because both estimators are valid in all cases, there is little to choose between them. The IVQR always has smaller bias than the MM-QR, but often has larger SE. As a result, the MM-QR generally has smaller MSE than the IVQR, but in general the performance of the estimators is very evenly matched. From a robustness point of view, it is reassuring to verify that the MM-QR estimator performs well even when the errors have high skewness and kurtosis.

To investigate the quality of the inference based on an estimator the covariance matrix implied by Theorem 5, we used it to computed the coverage of 95% confidence intervals centered at the MM-QR estimates; these results are presented in Table 5. Overall, the estimated coverage is close to the nominal level, except for $\tau = 0.75$ in Case 2 where the coverage drops to about 90%.

[17]This estimator was implemented using a grid search with 20 equally-spaced points between $\pm\left(60/\sqrt{N}\right) \times 100\%$ of the true parameter.

Table 4: Bias and MSE results with $\tau = 0.25$

		$n = 200$		$n = 1000$		$n = 5000$	
λ		IVQR	MM-QR	IVQR	MM-QR	IVQR	MM-QR
				Case 1: $\mathcal{N}(0,1)$			
0.50	BIAS	0.093	0.118	0.017	0.023	0.002	0.004
	SE	0.612	0.629	0.271	0.253	0.124	0.111
	MSE	0.384	0.409	0.074	0.064	0.015	0.012
0.25	BIAS	0.040	0.044	0.007	0.008	0.000	0.001
	SE	0.455	0.400	0.200	0.174	0.090	0.077
	MSE	0.209	0.162	0.040	0.030	0.008	0.006
				Case 2: $\chi^2_{(5)}$			
0.50	BIAS	0.071	0.079	0.014	0.019	0.002	0.004
	SE	0.440	0.393	0.195	0.159	0.089	0.071
	MSE	0.199	0.161	0.038	0.026	0.008	0.005
0.25	BIAS	0.033	0.040	0.007	0.008	0.001	0.002
	SE	0.322	0.268	0.142	0.117	0.065	0.053
	MSE	0.105	0.074	0.020	0.014	0.004	0.003
				Case 3: $t_{(5)}$			
0.50	BIAS	0.057	0.071	0.015	0.020	0.004	0.005
	SE	0.566	0.594	0.252	0.258	0.111	0.113
	MSE	0.323	0.358	0.064	0.067	0.012	0.013
0.25	BIAS	0.021	0.033	0.005	0.006	0.001	0.002
	SE	0.414	0.408	0.186	0.182	0.082	0.081
	MSE	0.172	0.167	0.034	0.033	0.007	0.007

Table 5: Coverage rates of 95% confidence intervals

	$n = 200$		$n = 1000$		$n = 5000$	
	$\lambda = 0.50$	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 0.25$	$\lambda = 0.50$	$\lambda = 0.25$
			Case 1: $\mathcal{N}(0,1)$			
$\tau = .25$	0.9423	0.9337	0.9406	0.9397	0.9413	0.9423
$\tau = .75$	0.9434	0.9367	0.9430	0.9400	0.9429	0.9420
			Case 2: $\chi^2_{(5)}$			
$\tau = .25$	0.9628	0.9485	0.9634	0.9532	0.9620	0.9522
$\tau = .75$	0.9151	0.9160	0.9211	0.9187	0.9020	0.9162
			Case 3: $t_{(5)}$			
$\tau = .25$	0.9386	0.9351	0.9378	0.9358	0.9404	0.9405
$\tau = .75$	0.9415	0.9375	0.9428	0.9394	0.9341	0.9346

6. ILLUSTRATIVE APPLICATIONS

In this section we present two examples illustrating that the proposed methods lead to results that are comparable to those obtained with approaches that are computationally much more demanding. To facilitate the comparison of our results with those in the extant literature, we only consider linear specifications of the conditional quantiles.

6.1. The determinants of government surpluses

Persson and Tabellini (2003) study the economic effects of constitutional reforms by looking at the relation between measures of economic performance and countries' economic, social, cultural, and political characteristics. For this illustration we focus on the determinants of the budget surplus (see Persson and Tabellini, 2003, Ch. 3).

Persson and Tabellini (2003) use data from 1960 to 1998 for 58 countries to estimate the relation between the surplus of the central government in percent of GDP (denoted SPL) and the following set of country characteristics: POLITY, the measure of the quality of democracy developed by Eckstein and Gurr (1975);[18] LYP, the log of real per capita income; TRADE, the sum of exports and imports of goods and services in percent of GDP; P1564, the percentage of the population between 15 and 64 years of age; P65, the percentage of the population over the age of 65; LSPL, one-year lag of SPL; OILIM, oil prices in US dollars times a dummy variable equal to 1 if the country is a net importer of oil; OILEX, oil prices in US dollars times a dummy variable equal to 1 if the country is a net exporter of oil; and YGAP, the output gap.[19] See Persson and Tabellini (2003) for full details on the sources and definition of variables used.

[18] Higher values of the index indicate worse democracies.

[19] The assumption that the regressors are not serially correlated is likely to be violated in this dynamic model; other assumptions are also likely to be violated (e.g., the linearity of all quantiles). The purpose of these applications, however, is essentially to illustrate that the proposed methods deliver results that are comparable to those obtained will less restrictive, but much more demanding, approaches.

The first two rows in Table 6 display the estimates of the parameters in the location and scale functions, together with analytical standard errors in parenthesis and clustered standard errors (estimated by bootstrap resampling countries) in square brackets.[20] As noted above, we assumed that the scale function is linear so as to preserve the linearity of the quantiles and facilitate the comparison with the estimates obtained with other methods. The results in rows 1 and 2 show that POLITY has effects with opposite signs on the location and scale,[21] suggesting that increasing the quality of the democracies reduces de average surplus, but also increases the dispersion of observed surpluses.

Rows 3 to 5 of Table 6 report the quantile regression estimates obtained with the MM-QR estimator presented in Section 3, and rows 6 to 8 display the jackknife-corrected MM-QR estimates, which in this case are very similar to the original ones. Again, we report in parenthesis the analytical standard errors based on an estimator of the covariance matrix given by Theorem 3, and in square brackets standard errors obtained by bootstrap (resampling by country), and note that in this example both sets of standard errors are very similar.

For comparison, rows 9 to 11 display quantile regression estimates of the same model using the method proposed by Canay (2011), which treats the fixed effects as location shifts. Because the model contains a lagged dependent variable, we also estimated the model using the method proposed by Galvão (2011).[22] To allow the fixed effects to differ across quantiles, Galvão's (2011) estimator was applied to each quantile at the time; these results are presented in rows 12 to 14. For the Canay (2011) and Galvão (2011) estimators we report only bootstrap (resampling by country) standard errors.

[20] The estimates in the first row match those reported by Persson and Tabellini (2003) in column 4 of their Table 3.4. Notice, however, that the original data used in the book contained some mistakes; the correct results and the data are available at Guido Tabellini's web-page: http://faculty.unibocconi.eu/guidotabellini/.

[21] Similar effects are observed for LSPL and OILIM

[22] We implemented the estimator using a grid search between 0.30 and 0.95 in steps of 0.01, and using the lag of LSPL as an instrument for it.

Table 6: The determinants of government surpluses

	POLITY	LYP	TRADE	P1564	PP65	LSPL	OILIM	OILEX	YGAP
OLS									
Location	0.12	−0.72	0.03	0.12	0.03	0.69	−0.05	−0.01	0.01
	(0.05)	(0.47)	(0.01)	(0.03)	(0.07)	(0.03)	(0.01)	(0.02)	(0.02)
	[0.05]	[0.50]	[0.01]	[0.03]	[0.08]	[0.04]	[0.01]	[0.02]	[0.02]
Scale	−0.10	−0.62	0.00	0.04	0.09	−0.08	0.01	0.02	−0.00
	(0.05)	(0.81)	(0.01)	(0.03)	(0.07)	(0.03)	(0.00)	(0.01)	(0.01)
	[0.05]	[0.76]	[0.01]	[0.03]	[0.07]	[0.03]	[0.00]	[0.01]	[0.01]
MM-QR									
$\tau = .25$	0.19	−0.24	0.03	0.09	−0.04	0.76	−0.06	−0.02	0.01
	(0.06)	(0.74)	(0.01)	(0.05)	(0.10)	(0.05)	(0.01)	(0.03)	(0.04)
	[0.07]	[0.74]	[0.01]	[0.04]	[0.11]	[0.03]	[0.01]	[0.03]	[0.03]
$\tau = .50$	0.11	−0.76	0.03	0.12	0.03	0.68	−0.05	−0.00	0.01
	(0.05)	(0.53)	(0.01)	(0.03)	(0.07)	(0.03)	(0.01)	(0.02)	(0.03)
	[0.05]	[0.51]	[0.01]	[0.03]	[0.08]	[0.04]	[0.01]	[0.02]	[0.02]
$\tau = .75$	0.03	−1.26	0.03	0.15	0.10	0.62	−0.04	0.01	0.01
	(0.06)	(0.65)	(0.01)	(0.04)	(0.08)	(0.04)	(0.01)	(0.03)	(0.03)
	[0.04]	[0.87]	[0.01]	[0.04]	[0.08]	[0.05]	[0.01]	[0.03]	[0.02]
MM-QR with jackknife bias correction									
$\tau = .25$	0.20	0.10	0.03	0.09	−0.06	0.76	−0.06	−0.02	0.01
	(0.06)	(0.74)	(0.01)	(0.05)	(0.10)	(0.05)	(0.01)	(0.03)	(0.04)
	[0.07]	[0.74]	[0.01]	[0.04]	[0.11]	[0.03]	[0.01]	[0.03]	[0.03]
$\tau = .50$	0.11	−0.79	0.03	0.12	0.04	0.68	−0.05	−0.00	0.01
	(0.05)	(0.53)	(0.01)	(0.03)	(0.07)	(0.03)	(0.01)	(0.02)	(0.03)
	[0.05]	[0.51]	[0.01]	[0.03]	[0.08]	[0.04]	[0.01]	[0.02]	[0.02]
$\tau = .75$	0.02	−1.65	0.04	0.16	0.13	0.61	−0.03	0.01	0.01
	(0.06)	(0.65)	(0.01)	(0.04)	(0.08)	(0.04)	(0.01)	(0.03)	(0.03)
	[0.04]	[0.87]	[0.01]	[0.04]	[0.08]	[0.05]	[0.01]	[0.03]	[0.02]
Canay									
$\tau = .25$	0.13	−0.84	0.03	0.15	0.05	0.74	−0.06	−0.02	0.05
	[0.06]	[0.51]	[0.01]	[0.03]	[0.08]	[0.02]	[0.01]	[0.03]	[0.02]
$\tau = .50$	0.10	−0.67	0.03	0.11	0.04	0.70	−0.05	−0.01	0.03
	[0.05]	[0.49]	[0.01]	[0.03]	[0.08]	[0.03]	[0.01]	[0.03]	[0.03]
$\tau = .75$	0.11	−0.76	0.03	0.10	0.04	0.65	−0.03	0.03	0.01
	[0.05]	[0.51]	[0.01]	[0.03]	[0.08]	[0.04]	[0.01]	[0.03]	[0.02]
Galvão									
$\tau = .25$	0.15	−0.50	0.03	0.12	0.02	0.76	−0.06	−0.01	0.04
	[0.07]	[0.61]	[0.01]	[0.05]	[0.08]	[0.05]	[0.01]	[0.04]	[0.03]
$\tau = .50$	0.05	0.01	0.02	0.08	−0.01	0.71	−0.05	−0.00	0.01
	[0.04]	[0.32]	[0.01]	[0.02]	[0.05]	[0.04]	[0.01]	[0.03]	[0.03]
$\tau = .75$	0.06	−0.30	0.02	0.10	0.05	0.65	−0.03	0.00	0.01
	[0.05]	[0.59]	[0.01]	[0.04]	[0.08]	[0.06]	[0.01]	[0.04]	[0.03]

The dependent variable is SPL; all regressions include country fixed effects. Unbalanced panel with 58 countries and 1659 observations. Analytical standard errors are in parenthesis and clustered standard errors (estimated by bootstrap resampling countries) are in square brackets.

For most variables, all quantile regression estimators lead to similar conclusions in terms of the magnitude and significance of the estimates. For example, all methods lead to very similar estimates of the coefficient on LSPL, the lagged dependent variable. However, there are also some very important differences between the results obtained with the different methods, especially between the results obtained with Canay's (2011) estimator and the results of less restrictive Galvão (2011) and MM-QR estimators.

Indeed, the results obtained with the Galvão and MM-QR estimators suggest that the effect of the quality of the democracy is very heterogeneous, being large for countries whose budget surplus is low relatively to that of countries with similar characteristics, and negligible for countries with high budget surpluses relatively to that of countries with similar characteristics. This pattern is in line with what could be expected from the estimates of the location and scale functions, and it is particularly clear in the results of the MM-QR estimator, for which the difference between the estimates for $\tau = 0.25$ and $\tau = 0.75$ is statistically significant at the 5% level. This finding contrasts sharply with the results obtained with Canay's (2011) estimator, which suggest that the effect of the quality of the democracy is essentially the same across the three quartiles, a result that does not accord with the estimates of the parameters in the scale function.

The time-series in this panel vary in length from 2 to 38 observations and therefore it is proper to be concerned with the validity of estimators that require large T. To check the robustness of the results, the estimations were repeated using only data for the 55 countries for which there are at least 10 observations; this reduces the total sample size to 1640. The results obtained with all estimators were remarkably insensitive to dropping the shorter series, and essentially the same estimates were obtained with the two samples.

This data set is reasonably small and therefore all estimators are somewhat imprecise. An example of the challenges posed by these data is that the three quartiles estimated using Galvão's method cross in 14 occasions. In these cases, if valid, the additional structure imposed by the MM-QR estimator can be helpful. Overall, however, we find that in this particular application, the results obtained with Galvão's (2011) method are qualitatively similar to those obtained with the much simpler MM-QR estimator.

6.2. Returns to training

Chernozhukov and Hansen (2008) use the data studied by Abadie, Angrist, and Imbens (2002) to illustrate the application of their instrumental variable quantile regression (IVQR) estimator. Here we use the same data to illustrate the application of the MM-QR estimator in a situation where one of the explanatory variables of the model is endogenous.

Briefly, these data were obtained from a randomized experiment performed under the Job Training Partnership Act in which individuals were randomly assigned the offer of training, but had the option to reject it. Because only 60% of those offered training accepted the offer, the actual training is self-selected but the randomly assigned offer provides a credible instrument for it.

The data used by Chernozhukov and Hansen (2008) contains information on 5102 adult males. Besides details on training assignment and actual training status, the data contains information on earnings and on a number of individual characteristics such as age, education, and ethnic background. Further details on the data are provided in Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2008).

Table 7 reports different estimates of the returns to training at a range of conditional quantiles, and the corresponding analytical standard errors obtained from suitable estimates of the covariance matrix. As in Chernozhukov and Hansen (2008), for brevity we do not report the estimates of the parameters associated with the controls.

The first row of Table 7 reports the estimates of the returns to training obtained with Koenker and Bassett's (1978) estimator that ignores the possible endogeneity of the treatment status; these estimates are all positive and statistically and economically significant, suggesting that the training had a strong positive impact across the conditional distribution, especially in its center and upper tail.

Table 7: Returns to training at different quantiles

	$\tau = .15$	$\tau = .25$	$\tau = .50$	$\tau = .75$	$\tau = .85$
QR	1187	2510	4420	4678	4807
	(209)	(360)	(596)	(901)	(991)
IVQR	–200	500	300	2700	3200
	(630)	(708)	(964)	(1510)	(1616)
MM-QR	211	389	1008	1972	2575
	(810)	(782)	(855)	(1327)	(1713)

5102 observations; analytical standard errors in square brackets.

This contrasts with the results obtained using Chernozhukov and Hansen's (2008) estimator, where actual training status is instrumented with the assignment indicator.[23] Indeed, the results in the second row of Table 7 suggest that the training only had an economically and statistically significant impact on the extreme upper tail of the conditional distribution.

The results obtained with the MM-QR, in which the actual training status is again instrumented with the assignment indicator, paint a similar picture. Indeed, the effect of the treatment status variable is positive but not statistically significant at the 10% level both in the location and in the scale functions, suggesting that the training is unlikely to have had a significant impact on the lower tail of the distribution and, at best, may have had some impact on the upper tail.[24] The third row of Table 7 reports the MM-QR estimates of returns to training at a range of quantiles, and the corresponding standard errors obtained from an estimator of the covariance matrix implied by Theorem 5. These results confirm that the effect of the training in the lower tail of the conditional distribution was neither statistically nor economically significant. This is in line with the IVQR results and contrasts with the results that ignore the endogeneity of the treatment indicator. The MM-QR estimates for the impact of the training in the upper

[23] The estimator was implemented as in Chernozhukov and Hansen (2008); the reported standard errors are obtained from the same article.

[24] The estimates of the training parameter in the location and scale functions are, respectively, 1331 (p-value: 0.11) and 956 (p-value: 0.12). Notice that if the location-scale model is adequate, the conditional mean will be a conditional quantile and the slope parameters will be smaller than 1331 in the quantiles below the mean, and lager for the quantiles above.

tail are sizable, but never statistically significant, even at the 10% level. Considering the precision of the estimates, however, the MM-QR and IVQR results are reasonably close and effectively lead to the same conclusion:[25] allowing for the possible endogeneity of the treatment status we find that, if anything, the training only had an impact on the upper tail of the conditional distribution.

In these linear models, the validity of the MM-QR depends on assumptions that are stronger than those required by the IVQR but, when these assumptions are valid, the MM-QR has some potential advantages. For example, in this sample, the five structural quantile functions estimated by IVQR cross more than 200 times, whereas the MM-QR estimator leads to estimates of these functions that do not cross.[26] Imposing this restriction, which is necessarily true, may result in efficiency gains and improved small-sample behavior, as documented by Zhao (2000).

7. CONCLUSIONS

In a conditional location-scale model, the information provided by the conditional mean and the conditional scale function is equivalent to the information provided by regression quantiles in the sense that these functions completely characterize how the regressors affect the conditional distribution. This is the result we use to estimate quantiles from estimates of the conditional mean and of the conditional scale function. Our approach is more restrictive than the traditional quantile regression, but we believe that the additional structure we impose can be useful in many applied settings. In particular, our approach provides an easy way to estimate regression quantiles in situations where using the traditional approach that is difficult or impossible.

[25] We note that our estimates are even closer to the ones obtained using the fully automated plug-in estimator of Kaplan and Sun (2017).

[26] It is possible to combine the IVQR with the method proposed by Chernozhukov, Fernández-Val, and Galichon (2010) to obtain structural quantile functions that do not cross.

The two very different applications we present illustrate that our method leads essentially to the same conclusions that are obtained with methods that are computationally much more demanding. This suggests that the proposed estimator can, at least, be useful in an exploratory phase, for example to provide starting values for other methods and to guide in the choice of the limits of the gird searches used in the Chernozhukov and Hansen (2008) and Galvão (2011) estimators.

Even when the effects of the regressors on the distribution of interest are not limited to their effects on the location and scale functions, i.e., when the location-scale model is inadequate, making a serious effort to model the heteroskedasticity can still be useful in applied work. Heteroskedasticity is often viewed as a nuisance, or interesting only inasmuch as knowledge of it can be used to improve the estimation of the conditional mean (see, e.g., Leamer, 2010, and Romano and Wolf, 2017).[27] However, the specification and estimation of the scale function is a simple and convenient way of gaining information on how the regressors affect features of the conditional distribution of interest other than its central tendency. When the location-scale model not appropriate, the information that can be obtained from the location and scale functions is not as rich as that provided by conditional quantiles, but may be interesting in itself, especially when estimation of conditional quantiles is not practical.

There are a number of aspects of the proposed approach that would be interesting to investigate. In the present paper we do not develop tests for the assumption that the location-scale model is adequate in the sense that the effects of the regressors on the distribution of interest are limited to their effects on the location and scale functions. In Section 2 we suggested that such tests can be constructed as tests for overidentifying restrictions, but it may be possible to develop simpler regression-based procedures. Also, following Hahn and Newey (2004) and most of the ensuing literature (e.g., Galvão and Kato, 2018), we assumed that the data are independent across i and t, and it would be

[27] Of course, heteroskedasticity can also be of interest in itself; the literature on ARCH/GARCH models is a leading example of that (see, e.g., Engle, 2001).

interesting to study the conditions under which it is possible to relax this assumption. Additionally, it would be interesting to extend our results to the case where the models have both individual and time effects; we are not aware of any quantile regression estimator that allows the inclusion of two sets of fixed effects but this problem should be tractable using our approach. Finally, it would naturally be interesting to see if in other applications the results obtained with the proposed method are also similar to those obtained with computationally more demanding estimators, as was the case in the applications we considered.

APPENDIX

A1: Assumptions

The results in the paper were derived under the following assumptions.

(P): On the parameter space

(1) $(\alpha_i, \delta_i)_{i=1}^n \in \Theta_1$, $(\beta, \gamma) \in \Theta_2$, where Θ_1 and Θ_2 are compact subsets of $\mathbb{R}^{2n}$ and $\mathbb{R}^{2k}$, respectively.

(2) The true parameter values are interior points of Θ_1 and Θ_2.

(3) Let F_U be the c.d.f. of U satisfying (U1) below and F_U^{-1} its inverse. $\tau \in \mathcal{T} = (\epsilon, 1-\epsilon)$, for some $\epsilon > 0$. The interval $(\lim_{\tau \searrow \epsilon} q(\tau)\,;\, \lim_{\tau \nearrow (1-\epsilon)} q(\tau))$ is bounded.

(U): On the error term

(1) The random variables U_{it} are *i.i.d.* (across i and t) and independent of X_{it} and Z_{it}.

(2) The random variables U_{it} have a continuous density function f_U and $f_U(u) > \zeta > 0$, $\forall u \in \text{supp}(U)$.[28]

[28] Assumption (U2) implies that the c.d.f. F_U is strictly monotone and therefore that the quantiles $q(\tau), \tau \in \mathcal{T}$ are unique.

(3) $E|U|^{2+\nu} < \infty$ for some $\nu > 0$.[29]

(XZ): On the regressors

(1) The sequence of random k-vectors $\{X_{it}\}$ is *i.i.d.* for any fixed i and independent across t.

(2) Z_{it} is a random k-vector defined by $Z_{itl} = \mathcal{Z}_l(X_{it})$, for $l = 1, \ldots, k$, where $\mathcal{Z}_l : \mathbb{R} \to \mathbb{R}$ is a known function of class $\mathcal{C}^1$ for a.e.-X. (Z_{itl} denotes the l-th coordinate of the vector Z_{it}.)

(3) There exists a $\xi > 0$ such that $P\{\inf_{i,t}(\delta_i + Z'_{it}\gamma) > \xi\} = 1$.

(4) $\max_{i \leq n} E|X_{i1l}|^{2+\nu} < K < \infty$ for some K and $\nu > 0$, for $l = 1, \ldots, k_1$. (X_{i1l} denotes the l-th coordinate of the vector X_{i1}.)

(5) $\max_{i \leq n} E|Z_{i1l}|^{4+\nu} < K < \infty$ for some K and $\nu > 0$, for $l = 1, \ldots, k_2$.

(6) $(1/n)\sum_i E[(X_{i1} - \bar{X}_i)(X_{i1} - \bar{X}_i)']$ is uniformly p.d. and has a constant limit Q_{XX}.

(7) $(1/n)\sum_i E[(Z_{i1} - \bar{Z}_i)(Z_{i1} - \bar{Z}_i)']$ is uniformly p.d. and has a constant limit Q_{ZZ}.

(8) $\max_{i \leq n} E|Z_{i1a}Z_{i1b}X_{i1c}|^{2+\nu} < K < \infty$ and $\max_{i \leq n} E|Z_{i1a}X_{i1c}X_{i1d}|^{2+\nu} < K < \infty$ for some K and $\nu > 0$, for $a, b = 1, \ldots, k_2$ and $c, d = 1, \ldots, k_1$.[30]

(9) The matrices $(1/n)\sum_i E[\sigma_{i1}^2(X_{i1} - \bar{X}_i)(X_{i1} - \bar{X}_i)']$ and $(1/n)\sum_i E[\sigma_{i1}^2(Z_{i1} - \bar{Z}_i)(X_{i1} - \bar{X}_i)']$ have constant limits denoted by P_{XX} and P_{XZ}, respectively.

(DC): On the regressors and instruments

(1) $E[|D_l|^{2+\nu}] < K < \infty$ for some K and $\nu > 0$, for $l = 1, \ldots, k_D$. (D_l denotes the l-th coordinate of the vector D.)

[29] Assumption (U3) implies that $E|V|^{2+\nu}$, $V = 2U[I\{U \geq 0\} - P\{U \geq 0\}] - 1$, is also finite.

[30] Applying Minkovski's inequality it easy to see that this assumption implies that the $(2+\nu)$-th absolute moments of $\sigma_{it}Z_{it}X'_{it}$ and $\sigma_{it}X_{it}X'_{it}$ exist and are uniformly bounded.

(2) $E[|C_{1_l}|^{4+\nu}] < K < \infty$ $(l = 1, \ldots, k_1)$ and $E[|C_{2_l}|^{2+\nu}] < K < \infty$ $(l = 1, \ldots, k_2)$ for some K and $\nu > 0$.

(3) $E[|\sigma'(X'\gamma)|^{2+\nu}] < K < \infty$ and $E[1/(|\sigma(X'\gamma)|^{2+\nu})] < K < \infty$.

(4) $E[CC']$ is non-singular.

(5) $E[(\sigma'/\sigma)|U|CX'] - (E[(1/\sigma)\,\mathrm{sign}(U)CX'])(E[(1/\sigma)CX'])^{-1}(E[(\sigma'/\sigma)UCX'])$ and $E[(1/\sigma)CX']$ are non-singular.

A2. Proofs

Proof (Theorem 1):

PART I: CONSISTENCY OF $\hat{\beta}$. The result is well know. For future reference write, which is possible under assumption (XZ6) for n and T large,

$$\hat{\beta} - \beta = \left(\frac{1}{nT} \sum_{it} \tilde{X}_{it} \tilde{X}_{it}' \right)^{-1} \frac{1}{nT} \sum_{it} \tilde{X}_{it} R_{it},$$

where $\sum_{it}$ is used as shorthand for $\sum_i \sum_t$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$, $\bar{X}_i = (1/T) \sum_t X_{it}$, and $\tilde{X}_{it}$ converges to 0 in L_2. It is also well know that under our assumptions the consistency also holds when $n \to \infty$ for fixed T, or $T \to \infty$ for fixed n.

PART II: CONSISTENCY OF $\hat{\gamma}$. To simplify notation put $\hat{\theta}_{it} = \theta_{it}(\hat{R}_{it}, \hat{\eta}) = 2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \hat{\eta})$ and $\theta_{it} = \theta_{it}(\hat{R}_{it}, \eta)$. Now notice that

$$\hat{\delta}_i - \delta_i = \frac{1}{T} \sum_t (\hat{\theta}_{it} - \sigma_{it}) - \bar{Z}_i{}'(\hat{\gamma} - \gamma)$$

and

$$\hat{R}_{it} = R_{it} - \bar{R}_{i,T} - \tilde{X}_{it}'(\hat{\beta} - \beta),$$

where $\bar{R}_{i,T} = (1/T)\sum_t R_{it}, i = 1, \ldots, n$. Consequently, defining $\tilde{Z}_{it} = Z_{it} - \bar{Z}_i$, $\bar{Z}_i = (1/T)\sum_t Z_{it}$, the concentrated estimation equation for γ can be written as

$$\begin{aligned}\left(\frac{1}{nT}\sum_{it}\tilde{Z}_{it}\tilde{Z}_{it}'\right)(\hat{\gamma}-\gamma) &= \frac{1}{nT}\sum_{it} Z_{it}\left[(\hat{\theta}_{it}-\sigma_{it}) - \frac{1}{T}\sum_t(\hat{\theta}_{it}-\sigma_{it})\right] \\ &= \frac{1}{nT}\sum_{it}\tilde{Z}_{it}(\hat{\theta}_{it}-\sigma_{it}) \\ &= \frac{1}{nT}\sum_{it}\tilde{Z}_{it}(\theta_{it}-\sigma_{it}) + o_P(1).\end{aligned}$$

The $o_P(1), (n,T) \to \infty$, remainder is justified by the fact that (letting $||\cdot||$ denote the L_2-norm),

$$\begin{aligned}\left\|(\hat{\eta}-\eta)\frac{1}{nT}\sum_{it}\tilde{Z}_{it}\hat{R}_{it}\right\| &\leq 2||\hat{\beta}-\beta||\,||Q_{XZ}|| \\ &\quad + 2\left\|\frac{1}{nT}\sum_{it}\tilde{Z}_{it}(R_{it}-\bar{R}_{i,T})\right\| \\ &\leq o(1) + 2\left\|\frac{1}{nT}\sum_{it}\tilde{Z}_{it}R_{it}\right\|,\end{aligned}$$

and the second term on the right-hand side is also $o(1)$ (actually $O(1/nT)$).

For what follows we need to introduce extra notation. Rewrite θ_{it} as

$$\begin{aligned}\theta_{it} &= 2(R_{it}-\bar{R}_{i,T}-\tilde{X}_{it}'(\hat{\beta}-\beta))[I\{R_{it}-\bar{R}_{i,T}-\tilde{X}_{it}'(\hat{\beta}-\beta)>0\}-\eta] \\ &= \theta_{it}(\bar{R}_{i,T},\hat{\beta}-\beta),\end{aligned}$$

and let,

$$M_{it} = M_{it}(\bar{R}_{i,T},\hat{\beta}-\beta) = \tilde{Z}_{it}[\theta_{it}(\bar{R}_{i,T},\hat{\beta}-\beta)-\sigma_{it}],$$

$$M_{n,t} = M_{n,t}(\bar{R}_{i,T},\hat{\beta}-\beta) = \frac{1}{n}\sum_i M_{it}, \qquad M_t^0 = E[M_{n,t}(\bar{R}_{i,T},0)].$$

Using this notation one may write,

$$\left(\frac{1}{nT}\sum_{it}\tilde{Z}_{it}\tilde{Z}_{it}'\right)(\hat{\gamma}-\gamma) = \frac{1}{T}\sum_t M_{n,t}.$$

The proof proceeds by establishing two claims:

Claim 1: $(1/T)\sum_t(M_{n,t}-M_t^0) \xrightarrow{P} 0$ *as* $(n,T)\to\infty$;

Claim 2: $(1/T)\sum_t M_t^0 = o(1)$ as $T\to\infty$.

These claims prove that $(1/T)\sum_t M_{n,t}$ and, thus, $(\hat{\gamma}-\gamma)$ is $o_P(1)$ as $(n,T)\to\infty$.

Proof of Claim 1:

$$||M_{n,t}-M_t^0|| \le ||M_{n,t}(\bar{R}_{i,T},\hat{\beta}-\beta)-M_{n,t}(\bar{R}_{i,T},0)|| + ||M_{n,t}(\bar{R}_{i,T},0)-M_t^0||$$

Since $f(v)=vI\{v>0\}$ is Lipschitz, $(||f(v-m)-f(v)||\le 2||m||)$, the first term is bounded by

$$\begin{aligned}&\Bigg\| 2\frac{1}{n}\sum_i \tilde{Z}_{it}\left[(R_{it}-\bar{R}_{i,T}-\tilde{X}_{it}'(\hat{\beta}-\beta))\,I\{R_{it}-\bar{R}_{i,T}>\tilde{X}_{it}'(\hat{\beta}-\beta)\}\right.\\ &\qquad \left.-\,(R_{it}-\bar{R}_{i,T})\,I\{R_{it}-\bar{R}_{i,T}>0\}\right]\Bigg\|\\ &\le 4||(\hat{\beta}-\beta)||\frac{1}{n}\sum_i ||\tilde{Z}_{it}||\,||\tilde{X}_{it}||.\end{aligned}$$

This term is $o(1)$ uniformly in t because $\hat{\beta}$ is consistent in L_2 and Z_{it} and X_{it} have, by assumption, uniformly bounded second moments. Also, $(1/T)\sum_t ||M_{n,t}(\bar{R}_{i,T},0)-M_t^0||$ converges to 0 in L_2 since it has mean 0 and a variance that, owing to the independence over i of U_{it}, is of order $O(1/n^{1/2})$.

Proof of Claim 2: A Taylor series expansion around $\bar{R}_{i,T}=0$ yields,

$$\frac{1}{T}\sum_t M_t^0 = \frac{1}{T}\sum_t E[M_{n,t}(0,0)] + \xi_{n,T}.$$

The leading term is 0 since

$$E\left[M_{it}(0,0)\right] = E\{\sigma_{it}\tilde{Z}_{it}[2U_{it}(I\{U_{it}>0\}-\eta)-1]\}$$

and $E\left[U\right]=0$ and $E\left[|U|\right]=1$ imply that $E[U_{it}(I\{U_{it}>0\}-\eta)] = E[U_{it}(I\{U_{it}>0\}] = 1/2$.

The remainder is

$$\xi_{n,T} = \frac{1}{nT}\sum_{it}\mu_{it}(\bar{R}_{i,T}^{\star})\bar{R}_{i,T},$$

where $||\bar{R}^{\star}_{i,T}|| \leq ||\bar{R}_{i,T}||$ and

$$\begin{aligned}\mu_{it}(\bar{R}^{\star}_{i,T}) &= \left.\frac{\partial E\left[M_{it}(y,0)\right]}{\partial y}\right|_{y=\bar{R}^{\star}_{i,T}} \\ &= -2\tilde{Z}_{it}\{(R_{it}-\bar{R}^{\star}_{i,T})f_{R_{it}}(\bar{R}^{\star}_{i,T}) + E[I\{R_{it}-\bar{R}^{\star}_{i,T}) > 0\} - \eta]\},\end{aligned}$$

where $f_{R_{it}}(\cdot)$ is the density of R_{it}, that is f_U/σ_{it}, and all expectations are conditional on the regressors. Under our assumptions on the parameter space and on the moments of U and Z_{it}, there exists a $K < \infty$ such that, $||R_{it}|| \leq [\|\delta_i\| + \|\gamma\| \; \|Z_{it}\|] \times \|U_{it}\| \leq K$. That is, R_{it}, and hence $\bar{R}_{i,T}$ and $\bar{R}^{\star}_{i,T}$, are uniformly L_2-bounded. Since f_U/σ_{it} is continuous and σ_{it} is uniformly bounded away from 0, $f_{R_{it}}(\bar{R}^{\star}_{i,T})$ is uniformly bounded and, consequently, so is $\mu_{it}(\bar{R}^{\star}_{i,T})$. Therefore, for some finite K'

$$\begin{aligned}||\xi_{n,T}|| &\leq K'\left\|\frac{1}{nT}\sum_{it}\bar{R}_{i,T}\right\| \\ &\leq K'\frac{1}{n}\sum_{i}||\bar{R}_{i,T}|| \\ &\leq K'\frac{1}{n}\sum_{i}\left\{(1/T^2)E[U_{i1}^2]\sum_{t}\sigma_{it}^2\right\}^{1/2} \\ &\leq T^{-1/2}K'',\end{aligned}$$

for some $K'' \leq \infty$. This completes the proof of Part II.

Part III: Consistency of $\hat{q}(\tau)$

Let $\hat{q}$ solve $\min_q S_{n,T}(q) = (1/nT)\sum_{it}\rho_\tau(\hat{R}_{it} - q\hat{\sigma}_{it})$, with $\hat{\sigma}_{it} = \hat{\delta}_i + Z'_{it}\hat{\gamma}$. By well-known arguments, it suffices to show that

$$S_{n,T}(q) \xrightarrow{P} E[\rho_\tau(U_{it}-q)].$$

The compactness of the parameter space (or the convexity of ρ_τ) implies that the convergence is uniform in q.

The sample objective function can be written as,

$$S_{n,T}(q) = (1/nT)\sum_{it}\rho_\tau(R_{it} - q\sigma_{it} - h_{it,T}),$$

with

$$h_{it,T} = \bar{R}_{i,T} + \tilde{X}_{it}'(\hat{\beta} - \beta) + q\frac{1}{T}\sum_t(\hat{\theta}_{it} - \sigma_{it}) + q\tilde{Z}_{it}'(\hat{\gamma} - \gamma).$$

Since $||\rho_\tau(v - h) - \rho_\tau(v)|| \leq ||h||$,

$$\left\|(1/nT)\sum_{it}\rho_\tau(R_{it} - q\sigma_{it} - h_{it,T}) - (1/nT)\sum_{it}\rho_\tau(R_{it} - q\sigma_{it})\right\| \leq (1/nT)\sum_{it}||h_{it,T}||,$$

and previous results show that the right-hand side is $o(1)$. The proof is completed by noting that the LLN implies that

$$(1/nT)\sum_{it}\rho_\tau(R_{it} - q\sigma_{it}) \xrightarrow{P} E[\rho_\tau(U_{it} - q)]. \qquad \blacksquare$$

For simplicity, the proofs of the other theorems will be decomposed into a series of partial results (lemmata). Some are merely instrumental, others may be of interest on their own. For economy of space we will not refer to any of the assumption above in the statement of these results. In rest of the appendix we will use the following notation

$$\begin{aligned}
\Delta_{1i} = \Delta_{1i_{n,T}} &= \sqrt{T}(\hat{\alpha}_i - \alpha_i),\\
\Delta_2 = \Delta_{2_{n,T}} &= \sqrt{nT}(\hat{\beta} - \beta),\\
\Delta_{3i} = \Delta_{3i_{n,T}} &= \sqrt{T}(\hat{\delta}_i - \delta_i),\\
\Delta_4 = \Delta_{4_{n,T}} &= \sqrt{nT}(\hat{\gamma} - \gamma),\\
\Delta_5 = \Delta_{5_{n,T}} &= \sqrt{nT}(\hat{q} - q(\tau)).
\end{aligned}$$

Lemma 1 *If $n/T \to 0$ as $(n, T) \to \infty$,*

$$\max_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_i| = o_P(1).$$

Proof: Standard least squares results show that

$$\Delta_{2_{n,T}} = Q_{XX}^{-1}\frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}(X_{it} - \bar{X}_i)U_{it} + o_P(1)$$

where, as before, $\bar{X}_i = (1/T)\sum_t X_{it}$, and

$$\Delta_{1i_{n,T}} = -\frac{1}{\sqrt{n}}\bar{X}_i'\Delta_{2_{n,T}} + \frac{1}{\sqrt{T}}\sum_t\sigma_{it}U_{it} = \frac{1}{\sqrt{T}}\sum_t\sigma_{it}U_{it} + o_P(1).$$

For any n and T, $E[\hat{\alpha}_i - \alpha_i] = 0$, $V(\hat{\beta} - \beta) = O(1/nT)$, and

$$V(\hat{\alpha}_i - \alpha_i) = \bar{X}_i' V(\hat{\beta} - \beta)\bar{X}_i + \frac{E[U^2]}{T^2}\sum_t \sigma_{it}^2 = O(1/nT) + O(1/T) = O(1/T).$$

Consider now $\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_i|$.

$$\begin{aligned}
P\{\max_{1\leq i\leq n} |\hat{\alpha}_i - \alpha_i| > \epsilon\} &\leq \sum_i P\{|\hat{\alpha}_i - \alpha_i| > \epsilon\} \\
&\leq \frac{1}{\epsilon^2}\sum_i V(\hat{\alpha}_i - \alpha_i) \\
&\leq \frac{1}{T\epsilon^2}\left(\frac{1}{n}\sum_i \bar{X}_i' V(\Delta_{2_{n,T}})\bar{X}_i\right) + \frac{E[U^2]}{\epsilon^2}\frac{n}{T}\left(\frac{1}{nT}\sum_{it}\sigma_{it}^2\right) \\
&\leq O(1/T) + \frac{n}{T}O(1) = o(1) \text{ if } \quad n/T \to 0.
\end{aligned}$$

■

Lemma 2 *Let $\hat{R}_{it} = Y_{it} - \hat{\alpha}_i - X_{it}\hat{\beta}$ and $\eta = E(I\{U > 0\})$. Then, as $(n, T) \to \infty$ with $n = o(T)$,*

$$\frac{1}{\sqrt{T}}\sum_t [2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}] - \frac{1}{\sqrt{T}}\sum_t \sigma_{it}[2U_{it}(I\{U_{it} > 0\} - \eta) - 1] = o_P(1) \ (i = 1, \ldots, n),$$

and

$$\frac{1}{\sqrt{nT}}\sum_{it} Z_{it}[2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}] - \frac{1}{\sqrt{nT}}\sum_{it} Z_{it}\sigma_{it}[2U_{it}(I\{U_{it} > 0\} - \eta) - 1] = o_P(1).$$

Proof: Put,

$$L_{n,T}(X_{it}, \Delta) = (1/\sqrt{T})\Delta_{1i} + (1/\sqrt{nT})X_{it}'\Delta_2,$$

and

$$\begin{aligned}
M_{2_{n,T}}(\Delta) &= \frac{1}{\sqrt{nT}}\sum_{it} Z_{it}[2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}] \\
&= \frac{1}{\sqrt{nT}}\sum_{it} Z_{it}[(2\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta)) \\
&\quad \times (I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \eta) - \sigma_{it}]
\end{aligned}$$

($\Delta = ((\Delta_{1i})_1^n, \Delta_2)$), and

$$\widetilde{M}_{2_{n,T}}(\Delta) = M_{2_{n,T}}(\Delta) - E[M_{2_{n,T}}(\Delta)].$$

We will first prove the stochastic equicontinuity of the empirical process $\widetilde{M}_{2_{n,T}}(\cdot)$. The proof will follow Andrews (1994). The function

$$m(U, Z, X, \delta_i, \gamma, \Delta) = [2\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta)][I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \eta]$$

is of CV-type I with envelope

$$\sup_{\delta_i,\gamma,\Delta} m(U, Z, X, \delta_i, \gamma, \Delta) = c_1 + c_2|U| + c_3\|Z\||U| + c_4\|X\|$$

for some constants c_i. Pollard's entropy condition (Andrews, 1994, section 4.2) is satisfied if

$$\lim_{n\to\infty}(1/n)\sum_i (E[\|Z_{i1}\|^{2+\nu} + 1) \sup_{\delta_i,\gamma,\Delta} \|m(U, Z, X, \delta_i, \gamma, \Delta)\|^{2+\nu} < \infty.$$

It suffices that,

$$\lim_{n\to\infty}\Bigg\{E|U|^{2+\nu} + \frac{1}{n}\sum_i [E\|Z_{i1}\|^{4+\nu} + E|U|^{2+\nu}E\|X_{i1}\|^{2+\nu} + E|U|^{2+\nu}E\|Z_{i1}\|^{2+\nu} + E|U|^{2+\nu}E\|X_{i1}Z_{i1}\|^{2+\nu} + E\|X_{i1}\|^{2+\nu}]\Bigg\} < \infty.$$

Assumptions (U2) and (XZ 4, 5, and 8) yield the desired result and prove the stochastic equicontinuity of $\widetilde{M}_{2_{n,T}}(\cdot)$.

Stochastic equicontinuity and the fact that $\max_i |(1/\sqrt{T})\Delta_{1i}| = o_P(1)$ and $(1/\sqrt{nT})\Delta_2 = o_P(1)$ imply (Andrews, 1994, p. 2265) that

$$\widetilde{M}_{2_{n,T}}(\Delta) - \widetilde{M}_{n,T}(0) = o_P(1).$$

Consequently,

$$\begin{aligned} M_{2_{n,T}}(\Delta) &= E[M_{2_{n,T}}(\Delta)] + \widetilde{M}_{2_{n,T}}(0) + [\widetilde{M}_{2_{n,T}}(\Delta) - \widetilde{M}_{2_{n,T}}(0)] \\ &= E[M_{2_{n,T}}(\Delta)] + \widetilde{M}_{2_{n,T}}(0) + o_P(1) \\ &= E[M_{2_{n,T}}(\Delta)] + M_{2_{n,T}}(0), \end{aligned}$$

since $E[M_{2_{n,T}}(0)] = 0$.

The lemma is proved as a first-order Taylor series expansion of $E[M_{n,T}(\Delta)]$ around $\Delta = 0$ yields

$$E[M_{2_{n,T}}(\Delta)] = -E[I\{U_{it} > 0\} - \eta]\frac{1}{\sqrt{nT}}\sum_{it} E[L_{n,T}(X_{it}, \Delta)] = 0.$$

Now put

$$M_{1_{n,T}}(\Delta) = \frac{1}{\sqrt{T}}\sum_{t}[2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \eta) - \sigma_{it}].$$

The same arguments yield

$$M_{1_{n,T}}(\Delta) = M_{1_{n,T}}(0) + o_P(1).$$

■

Lemma 3 *Let,*

$$\hat{\eta} = \frac{1}{nT}\sum_{it} I\{\hat{R}_{it} > 0\}.$$

Then,

$$\sqrt{nT}(\hat{\eta} - \eta) = O_P(1) \quad \text{as} \quad (n, T) \to \infty \quad \text{with} \quad n = o(T).$$

Proof: Using the notation of lemma 2, let,

$$\widetilde{R}_{n,T}(U, X, \Delta) = \frac{1}{nT}\sum_{it} I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\} - \frac{1}{nT}\sum_{it} E[I\{\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta) > 0\}].$$

The process $\widetilde{R}_{n,T}(\cdot)$ satisfies trivially Pollard's entropy condition and so it is equicontinuous (see Andrews, 1994, p. 2273). Since $\max_i |(1/\sqrt{T})\Delta_{1i}| = o_P(1)$ and $(1/\sqrt{nT})\Delta_2 = o_P(1)$,

$$\widetilde{R}_{n,T}(\cdot, \Delta) = \widetilde{R}_{n,T}(\cdot, 0) = o_P(1)$$

since $\widetilde{R}_{n,T}(\cdot, 0) = o_P(1)$ by the law of large numbers. Now, a Taylor series expansion yields,

$$\sqrt{nT}(\hat{\eta} - \eta) = -\frac{f_U(0)}{\sqrt{nT}}\sum_{it}\frac{1}{\sigma_{it}}\left[\frac{1}{\sqrt{T}}\Delta_{1i} + \frac{1}{\sqrt{nT}}X_{it}'\Delta_2\right] + o_P(1),$$

which establishes the result for $(1/nT)\sum_{it}(1/\sigma_{it})X_{it} = O_P(1)$ and

$$\begin{aligned}\frac{1}{\sqrt{nT}}\sum_{it}\frac{1}{\sigma_{it}}\frac{1}{\sqrt{T}}\Delta_{1i} &= \frac{1}{\sqrt{nT}}\sum_{it}\frac{1}{\sigma_{it}}\frac{1}{\sqrt{T}}\left[-\frac{1}{\sqrt{n}}\bar{X}_i'\Delta_{2_{n,T}} + \frac{1}{\sqrt{T}}\sum_t \sigma_{it}U_{it}\right] \\ &= O_P(1) + \frac{1}{\sqrt{nT}}\sum_{it}\pi_{i,T}\sigma_{it}U_{it} \\ &= O_P(1),\end{aligned}$$

where the last equality follows from applying the central limit theorem (with $\pi_{i,T} = (1/T)\sum_t(1/\sigma_{it})$). ■

Proof (Theorem 2): The moment conditions defining the estimators of δ_i $(i = 1, \ldots, n)$ and γ are,

$$\frac{1}{\sqrt{T}}\sum_t \left\{[2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \hat{\eta}) - \sigma_{it}] - \frac{1}{\sqrt{T}}\Delta_{3i} - \frac{1}{\sqrt{nT}}Z_{it}'\Delta_4\right\} = 0$$
$$\frac{1}{\sqrt{nT}}\sum_{it} Z_{it}\left\{[2\hat{R}_{it}(I\{\hat{R}_{it} > 0\} - \hat{\eta}) - \sigma_{it}] - \frac{1}{\sqrt{T}}\Delta_{3i} - \frac{1}{\sqrt{nT}}Z_{it}'\Delta_4\right\} = 0,$$

which can be written as

$$G_n\begin{pmatrix}\Delta_{3i}\\ \Delta_4\end{pmatrix} = \begin{pmatrix}M_{1_{n,T}}(0)\\ M_{2_{n,T}}(0)\end{pmatrix} + (\hat{\eta} - \eta)\begin{pmatrix}0\\ \frac{1}{\sqrt{nT}}\sum_{it} Z_{it}[2\sigma_{it}U_{it} - L_{n,T}(X_{it}, \Delta_{1i}, \Delta_2)]\end{pmatrix}$$

with

$$G_n = \begin{pmatrix}1 & (1/\sqrt{n})\bar{Z}_i' \\ (1/\sqrt{n})\sum_i \bar{Z}_i & (1/nT)\sum_{it} Z_{it}Z_{it}'\end{pmatrix}.$$

where, as before, $\bar{Z}_i = (1/T)\sum_t Z_{it}$. Lemma 3 implies that the second term on the right-hand side is $o_P(1)$. Solving the system for Δ_4 gives,

$$Q_{ZZ}\Delta_4 = \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}(Z_{it} - \bar{Z}_i)[2U_{it}(I\{U_{it} > 0\} - \eta) - 1].$$

The central limit theorem establishes the desired result. ■

Lemma 4 *If $n/T \to 0$ as $(n, T) \to \infty$,*

$$\max_{1\le i\le n}|\hat{\delta}_i - \delta_i| = o_P(1).$$

Proof: The first equation of the system in the proof of Theorem 2 implies that (adopting the same notation)

$$\frac{1}{\sqrt{T}}\Delta_{3i} = \frac{1}{\sqrt{T}} M_{1_{n,T}}(0) - \frac{1}{\sqrt{nT}}\bar{Z}_i'\Delta_4.$$

For any $\epsilon > 0$,

$$\begin{aligned}
P\left\{\max_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\Delta_{3i}\right| > \epsilon\right\} &\leq \sum_i P\left\{\left|\frac{1}{\sqrt{T}}M_{1_{n,T}}(0)\right| > \frac{\epsilon}{2}\right\} \\
&\quad + \sum_i P\left\{\left|\frac{1}{\sqrt{nT}}\bar{Z}_i'\Delta_4\right| > \frac{\epsilon}{2}\right\} \\
&\leq \frac{2E[V^2]}{\epsilon^2}\frac{n}{T}\left(\frac{1}{nT}\sum_{it}E[\sigma_{it}^2]\right) \\
&\quad + \frac{1}{T}\left(\frac{1}{n}\sum_i E[\bar{Z}_i' E[\Delta_4\Delta_4']\bar{Z}_i]\right) \\
&= \frac{n}{T}O(1) + O(T^{-1}).
\end{aligned}$$

■

Proof (Theorem 3): Let $\psi_\tau(A) = -(I\{A \leq 0\} - \tau)$, $\Delta = ((\Delta_{1i})_1^n,\ \Delta_2,\ (\Delta_{3i})_1^n,\ \Delta_4, \Delta_5)$,

$$\begin{aligned}
\Psi_{n,T}(U,X,Z,\Delta) &= \frac{1}{\sqrt{nT}}\sum_{it}\hat{\sigma}_{it}\psi_\tau\left[\hat{R}_{it} - \hat{q}\hat{\sigma}_{it}\right] \\
&= \frac{1}{\sqrt{nT}}\sum_{it}\Big\{\left[\sigma_{it} + L_{nT}(Z_{it},(\Delta_{3i})_1^n,\ \Delta_4)\right]\psi_\tau\left(\sigma_{it}U_{it} - L_{nT}\left(X_{it},(\Delta_{1i})_1^n,\ \Delta_2\right)\right) \\
&\quad - \left(q(\tau) - \frac{1}{\sqrt{nT}}\Delta_5\right)\left(\sigma_{it} + L_{nT}\left(Z_{it},(\Delta_{3i})_1^n,\ \Delta_4\right)\right)\Big\} \\
&= o_P(1)
\end{aligned}$$

and

$$\widetilde{\Psi}_{n,T}(U,X,Z,\Delta) = \Psi_{n,T}(U,X,Z,\Delta) - E[\Psi_{n,T}(U,X,Z,\Delta)].$$

The boundedness of $\psi_\tau(\cdot)$ and the moment conditions suffice to yield the stochastic equicontinuity of $\widetilde{\Psi}_{n,T}(U,X,Z,\Delta)$. As, $\max_i|(1/\sqrt{T})\Delta_{1i}|$, $(1/\sqrt{nT})\Delta_2$, $\max_i|(1/\sqrt{T})\Delta_{3i}|$, $(1/\sqrt{nT})\Delta_4$, and $(1/\sqrt{nT})\Delta_5$ are all $o_P(1)$ as $(n,T)\to\infty$ with $n/T\to 0$,

$$\widetilde{\Psi}_{n,T}(U,X,Z,\Delta) - \widetilde{\Psi}_{n,T}(U,X,Z,0) = o_P(1).$$

Consequently (note that $E[\Psi_{n,T}(U,X,Z,0)] = 0$),

$$\Psi_{n,T}(U,X,Z,\Delta) = E[\Psi_{n,T}(U,X,Z,\Delta)] + \Psi_{n,T}(U,X,Z,0) + o_P(1).$$

The first term on the right-hand side can be approximated to the first order around $\Delta = 0$ by

$$E[\Psi_{n,T}(U,X,Z,\Delta)] = -f_U(q(\tau))\left\{\frac{1}{\sqrt{nT}}\sum_{it} L_{nT}(X_{it},(\Delta_{1i})_1^n,\ \Delta_2)\right.$$
$$\left. + q(\tau)L_{nT}\left(Z_{it},(\Delta_{3i})_1^n,\ \Delta_4\right) + \frac{1}{nT}\sum_{it}\sigma_{it}\Delta_5\right\}.$$

The second term

$$\Psi_{n,T}(U,X,Z,0) = \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}\psi_\tau(U_{it} - q(\tau))$$

is an asymptotically normal sequence. Putting the two terms together,

$$\sqrt{n}\bar{\Delta}_1 + \bar{\bar{X}}'\Delta_2 + q(\tau)[\sqrt{n}\bar{\Delta}_3 + \bar{\bar{Z}}'\Delta_4] = \frac{1}{f_U(q(\tau))}\frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}\psi_\tau(U_{it} - q(\tau)) + o_P(1),$$

with $\bar{\Delta}_1 = (1/n)\sum_i \Delta_{1i}$ and $\bar{\bar{X}} = (1/nT)\sum_{it} X_{it}$ (and likewise for $\bar{\Delta}_3$ and $\bar{\bar{Z}}$). Note that,

$$\sqrt{n}\bar{\Delta}_1 + \bar{\bar{X}}'\Delta_2 = \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}U_{it}$$

and

$$\sqrt{n}\bar{\Delta}_3 + \bar{\bar{Z}}'\Delta_4 = \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}V_{it}.$$

Consequently,

$$\mu_\sigma\Delta_5 = \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}\left[\frac{1}{f_U(q(\tau))}\psi_\tau(U_{it} - q(\tau)) + U_{it} + q(\tau))V_{it}\right].$$

Combining this result with the representation of Δ_4 in the proof of Theorem 2 and with the usual representation of the least squares estimator Δ_2 gives,

$$\begin{pmatrix} Q_{XX} & \mathcal{O} & \mathcal{O} \\ \mathcal{O} & Q_{ZZ} & \mathcal{O} \\ 0' & 0' & \mu_\sigma \end{pmatrix}\begin{pmatrix} \Delta_2 \\ \Delta_4 \\ \Delta_5 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}(X_{it} - \bar{X}_i)U_{it} \\ \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}(Z_{it} - \bar{Z}_i)V_{it} \\ \frac{1}{\sqrt{nT}}\sum_{it}\sigma_{it}W_{it} \end{pmatrix}.$$

where $\mathcal{O}$ and 0 denote a $k \times k$ matrix and k-vector of 0s, respectively. The result then follows from the central limit theorem and the delta-method. ■

Lemma 5 (Quantiles with Measurement error) *Consider three unobserved random variables (U, W_1, W_2) with joint density $f_{UW_1W_2}(\cdot)$ bounded away from zero and of class $\mathcal{C}^2$. Assume further that (W_1, W_2) have moments of order 3. The measurement error contaminated observations of U are given by*

$$U^* = \frac{U + W_2}{1 + W_1}.$$

Then, letting $q^0 = q^0(\tau) = F_U^{-1}(\tau)$ denote the τ-th marginal quantile of U,

$$\begin{aligned} P\{U^* \leq q^0\} = \tau &+ f_{U|W}(q^0|0)E[q^0 W_1 - W_2] \\ &+ \left(q^0 f^1_{U|W}(q^0|0) + (1/2)(q^0)^2 f^u_{U|W}(q^0|0)\right) E\left(W_1^2\right) \\ &+ \left((1/2) f^u_{U|W}(q^0|0) - f^2_{U|W}(q^0|0)\right) E\left(W_2^2\right) \\ &- \left(q^0 f^u_{U|W}(q^0|0) + f^1_{U|W}(q^0|0) - q^0 f^2_{U|W}(q^0|0)\right) E\left(W_1 W_2\right) + O(||W||^3) \end{aligned}$$

where $f_{U|W}$ is the conditional density of U given $W = (W_1, W_2)$, $f^u_{U|W} = \partial f_{U|W}(u|w)/\partial u$ and $f^j_{U|W} = \partial f_{U|W}(u|w)/\partial w_j$ (cf. Chesher, 2017).

Proof: The data identifies $q^1(\tau) = F^{-1}_{U^*}(\tau)$ and we want to approximate $q^0(\tau) = F_U^{-1}(\tau)$. Due to the contamination $E[I(U^* \leq q^0)] \neq \tau$, implying that,

$$E_{U|W}[I(U \leq q^0(1 + w_1) - w_2)|W_1 = w_1, W_2 = w_2] = F_{U|W}(q^0(1 + w_1) - w_2) \neq \tau.$$

Regard $F_{U|W}(q^0(1 + w_1) - w_2)$ as a function of q (say, $h(q)$) and expand it around q^0 given $W = (w_1, w_2)$,

$$h(q) = E_{U|W} I(U \leq q^0) + f_{U|W}(q^0|w)(q - q^0) + (1/2) f^u_{U|W}(q^0|w)(q - q^0)^2 + O(|q - q^0|^3)$$

with $q - q^0 = q^0 W_1 - W_2$. We now expand this partial derivative around $W = 0$. Notice that

$$f_{U|W}(q^0|w) = f_{U|W}(q^0|0) + \sum_{j=1}^{2} f^j_{U|W}(q^0|0) W_j + O(||W||^2)$$

and that,

$$f^u_{U|W}(q^0|w) = f^u_{U|W}(q^0|0) + O(||W||).$$

Plugging back in the expansion for $h(q)$ and taking expectations with respect to $W = (W_1, W_2)$ one gets,

$$\begin{aligned} E_W[F_{U|W}(q)] = \tau &+ f_{U|W}(q^0|0)E[q^0 W_1 - W_2] \\ &+ \left(q^0 f^1_{U|W}(q^0|0) + (1/2)(q^0)^2 f^u_{U|W}(0|0)\right) E\left(W_1^2\right) \\ &+ \left((1/2) f^u_{U|W}(0|0) - f^2_{U|W}(q^0|0)\right) E\left(W_2^2\right) \\ &- \left(q^0 f^u_{U|W}(q^0|0) + f^1_{U|W}(q^0|0) - q^0 f^2_{U|W}(q^0|0)\right) E\left(W_1 W_2\right). \end{aligned}$$

■

Proof (Theorem 4): The linear representation of the quantile estimator yields ($q \equiv q(\tau)$) as n approaches ∞,

$$\frac{1}{nT}\sum_{it} f_{U^*_{it}}(q)(\hat{q} - q) = \frac{1}{nT}\sum_{it}(\tau - E[I\{U^*_{it} \le q\}]) + o_P(1). \tag{10}$$

We will use lemma 5 with

$$W_1 = W_{1_{it}} = (1/\sigma_{it})\bar{S}_i - (1/\sigma_{it})\tilde{Z}'_{it}\gamma_T$$

and

$$W_2 = W_{2_{it}} = -\bar{R}_i/\sigma_{it},$$

to approximate $E[I\{U^*_{it} \le q\}]$. Under our assumptions W_1 and W_2 are independent over the i dimension and *i.i.d.* over t. To compute the (approximated) moments of W_1 and W_2 it is useful to establish some notation and basic results:

$$\begin{aligned} \bar{S}_{i,T} &= \frac{1}{T}\sum_a [2(R_{ia} - \bar{R}_{i,T})I\{R_{ia} - \bar{R}_{i,T} > 0\} - \sigma_{ia}] \\ &:= \frac{1}{T}\sum_a s(R_{ia}, \bar{R}_{i,T}) \\ &:= \frac{1}{T}\sum_a s_{ia} \end{aligned}$$

$$\gamma_T = \hat{\gamma} - \gamma = Q_{ZZ}^{-1} \frac{1}{T} \sum_a^T \frac{1}{n} \sum_j^n \tilde{Z}_{ja} s_{ja}.$$

For a (integrable) function $g(U_{it}, \bar{R}_i)$ let $E_{U|W}[g(U_{it}, \bar{R}_i)] = \int g(u,r) f_{U|W}(u|r)du$ denote the conditional expectation given $\bar{R}$. A power expansion yields,

$$\begin{aligned} E[s_{ia}] &= E_W E_{U|W} s(R_{ia}, \bar{R}_{i,T}) \\ &= \frac{f_{U|W}(0|0)}{\sigma_{ia}} E[\bar{R}_i^2] + O(E[|\bar{R}_i|^3]) \\ &= \frac{E[U^2] f_{U|W}(0|0)}{T} \frac{\overline{\sigma_i^2}}{\sigma_{ia}} + O(1/T^2), \end{aligned}$$

where $\overline{\sigma_i^2} = (1/T) \sum_a^T \sigma_{ia}^2$.[31] A similar expansion argument in powers of $\bar{R}_i$ yields,

$$\begin{aligned} E[s(R_{ia}, \bar{R}_i) s(R_{ib}, \bar{R}_i)] = 4\sigma_{ia}^2 E[(UI(U > 0) - 1)^2] \\ + 4E[I(U_{ia} > 0) I(U_{ib} > 0)] E[\bar{R}_i^2] + O(|\bar{R}_i|^3). \end{aligned}$$

Furthermore, for $i \neq j$

$$\begin{aligned} E[s(R_{ia}, \bar{R}_i) s(R_{jb}, \bar{R}_j)] &= E[s(R_{ia}, \bar{R}_i)] E[s(R_{jb}, \bar{R}_j)] \\ &= \frac{(E[U^2])^2 f_{U|W}^2(0|0)}{T^2} \frac{\overline{\sigma_i^2 \sigma_j^2}}{\sigma_{ia}\sigma_{ja}} \\ &= O(1/T^2). \end{aligned}$$

[31] It is possible to compute explicitly the marginal expectation with respect to U. Making $D_{ia,T} = (1/T) \sum_{t \neq a}^T \sigma_{it} U_{it}$,

$$\begin{aligned} E[s_{ia}] &= \sigma_{ia} E\left[2 \frac{T-1}{T} (U_{it} - D_{ia,T}) I\{U_{ia} > D_{ia,T}\} - 1 \right] \\ &= -\frac{\sigma_{ia}}{T} + \frac{\sigma_{ia}(T-1) f_U(0)}{T} \frac{E\left(U^2\right)}{(T-1)\sigma_{ia}^2} \overline{\sigma_{i,(-a)}^2} \\ &= -\frac{\sigma_{ia}}{T} + \frac{f_U(0) E(U^2)}{T} \frac{\overline{\sigma_{i,(-a)}^2}}{\sigma_{ia}} \end{aligned}$$

with $\overline{\sigma_{i,(-a)}^2} := \frac{1}{T-1} \sum_{a \neq t}^T \sigma_{ia}^2$.

From the expression for $E[s_{ia}]$ it follows directly that,

$$E[\gamma_T] = \frac{E[U^2]f_{U|W}(0|0)}{T}\frac{1}{nT}\sum_{it} Q_{ZZ}^{-1}\tilde{Z}_{it}\frac{\overline{\sigma_i^2}}{\sigma_{it}} + O(1/T^2)$$
$$:= \frac{E[U^2]f_{U|W}(0|0)}{T}\Gamma + O(1/T^2),$$

which gives the expression for the $O(1/T)$ term in the bias of $\hat{\gamma}$, that is, B_T^{γ}.

It is now easy to establish that,

$$E[W_{2_{it}}] = E[-\bar{R}_i/\sigma_{it}] = 0, \tag{11}$$
$$E[W_{1_{it}}] = \frac{E\left(U^2\right)f_{U|W}(0|0)}{T}\left(\frac{\pi_i\,\overline{\sigma_i^2}}{\sigma_{it}} - \tilde{Z}_{it}'\Gamma\right), \tag{12}$$

with $\pi_i = (1/T)\sum_a^T(1/\sigma_{ia})$.

Let us now turn to the second moments of W:

$$E[W_{2_{it}}^2] = \frac{E[\bar{R}_i^2]}{\sigma_{it}^2} = \frac{E[U^2]}{T}\frac{\overline{\sigma_i^2}}{\sigma_{it}^2}, \tag{13}$$

and

$$E[W_{1_{it}}^2] = \frac{E[\bar{S}_i^2]}{\sigma_{it}^2} + o(1)$$
$$= \frac{1}{T^2}\sum_a^T\sum_b^T E[s(R_{ia},\bar{R}_i)s(R_{ib},\bar{R}_i)]$$
$$= \frac{4}{T}\left\{E[(UI(U>0)-1)^2] + E\left(U^2\right)(1-F_{U|W}(0|0))^2\right\}\frac{\overline{\sigma_i^2}}{\sigma_{it}^2} + o(1/T), \tag{14}$$

where $F_{U|W}(0|0) = \int_{-\infty}^0 f_{U|W}(u|0)\,du$. The $o(1/T)$ as $n\to\infty$ remainder results from the fact that

$$E[\gamma_T\gamma_T'] = Q_{ZZ}^{-1}\left(\frac{1}{n^2T^2}\sum_j^n\sum_l^n\sum_a^T\sum_b^T \tilde{Z}_{ja}\tilde{Z}_{lb}'E\left[s_{ja}s_{lb}\right]\right)Q_{ZZ}^{-1}$$
$$= \frac{4E[U^2](1-F_{U|W}(0|0))^2}{nT^2}Q_{ZZ}^{-1}\left(\frac{1}{nT}\sum_j^n\sum_a^T\overline{\sigma_j^2}\tilde{Z}_{ja}\tilde{Z}_{ja}'\right)Q_{ZZ}^{-1}$$
$$= o(1/T^2),$$

and

$$E[\bar{S}_i\gamma_T] = Q_{ZZ}^{-1}\frac{1}{nT^2}\sum_j^n\sum_a^T\sum_b^T \tilde{Z}_{jb}E[s_{ia}s_{jb}] = Q_{ZZ}^{-1}\frac{1}{nT^2}\sum_a^T\sum_b^T \tilde{Z}_{ib}E[s_{ia}s_{ib}]$$

$$= Q_{ZZ}^{-1}\frac{4E[(UI(U>0)-1)^2]}{nT}\frac{1}{T}\sum_a^T \sigma_{ia}\tilde{Z}_{ia} = o(1/T).$$

Finally, noticing that

$$E[\bar{R}_i\gamma_T] = Q_{ZZ}^{-1}\frac{1}{nT}\sum_j^n\sum_a^T \tilde{Z}_{ja}E[s(R_{ja},\bar{R}_j)\bar{R}_i] = Q_{ZZ}^{-1}\frac{1}{nT}\sum_a^T \tilde{Z}_{ia}E[s(R_{ia},\bar{R}_i)\bar{R}_i]$$

$$= \frac{-E[U^2](1-F_{U|W}(0|0))}{nT}\overline{\sigma_i^2}Q_{ZZ}^{-1}\frac{1}{T}\sum_a^T \tilde{Z}_{ia} = 0,$$

one has

$$E[W_{1_{it}}W_{2_{it}}] = \frac{-1}{\sigma_{it}^2}E[\bar{R}_i\bar{S}_i] + \frac{1}{\sigma_{it}^2}E[\bar{R}_i\gamma_T] = \frac{-1}{\sigma_{it}^2}\frac{1}{T}\sum_a^T E[s(R_{ia},\bar{R}_i)\bar{R}_i] + 0$$

$$= \frac{E[U^2](1-F_{U|W}(0|0))}{T}\frac{\overline{\sigma_i^2}}{\sigma_{it}^2} + O(1/T^2). \tag{15}$$

Lemma 5 and (10) imply that

$$\frac{1}{nT}\sum_{it} f_{U_{it}^*}(q)(\hat{q}-q) = \frac{1}{nT}\sum_{it}(\tau - E[I\{U_{it}^* \le q\}]) + o_P(1)$$

$$= f_{U|W}(q|0)q\frac{1}{nT}\sum_{it}E[W_{1_{it}}]$$

$$+ (qf_{U|W}^1(q|0) + (1/2)q^2 f_{U|W}^u(q|0))\frac{1}{nT}\sum_{it}E[W_{1_{it}}^2]$$

$$+ ((1/2)f_{U|W}^u(q|0) - f_{U|W}^2(q|0))\frac{1}{nT}\sum_{it}E[W_{2_{it}}^2]$$

$$- (q^0 f_{U|W}^u(q^0|0) + f_{U|W}^1(q^0|0) - q^0 f_{U|W}^2(q^0|0))\frac{1}{nT}\sum_{it}E[W_{1_{it}}W_{2_{it}}]. \tag{16}$$

Note that $(1/nT)\sum_{i,t}E[W_{a_{it}}]$ and $(1/nT)\sum_{i,t}E[W_{a_{it}}W_{b_{it}}], a, b = 1, 2$, are all $O(1/T)$; see (11)-(15).

It remains to approximate $(1/nT)\sum_{it} f_{U^*_{it}}(q)$ in the left-hand side of (10) around $(1/nT)\sum_{it} f_{U_{it}}(q)$.

$$\begin{aligned}(1/nT)\sum_{it} f_{U^*_{it}}(q) &= (1/nT)\sum_{it} f_U\left(q(1+W_{1_{it}})\right)(1+W_{1_{it}}) \\ &= f_{U|W}(q) + O(\sum_{it} E[W_{1_{it}}]/nT) \\ &= f_{U|W}(q) + O(1/T). \qquad (17)\end{aligned}$$

Together, (16) and (17) imply that the bias of $(\hat{q}-q)$ has an expansion in powers of $1/T$ as required. ■

Proof (Theorem 5): Put $\Delta_1 = \sqrt{n}(\hat{\beta}-\beta), \Delta_2 = \sqrt{n}(\hat{\gamma}-\gamma), \Delta_3 = \sqrt{n}(\hat{q}-q(\tau))$. Standard GMM arguments (Newey and McFadden, 1994) or, for $\hat{q}$, arguments as in Theorem 1, prove the consistency of $(\hat{\beta},\hat{\gamma},\hat{q})$.

Let us start with the linear representation of Δ_3 conditional on root-n consistent estimators of β and γ,

$$\frac{Y-X'\hat{\beta}}{\sigma(X'\hat{\gamma})} - \hat{q} = (U-q(\tau)) - L_n(U,X,\Delta_1,\Delta_2) - \frac{1}{\sqrt{n}}\Delta_3 - K_n(U,X,,\Delta_1,\Delta_2),$$

where

$$L_n(U,X,\Delta_1,\Delta_2) = \frac{1}{\sigma}\frac{1}{\sqrt{n}}X'\Delta_1 + \frac{\sigma'}{\sigma}\frac{1}{\sqrt{n}}UX'\Delta_2,$$

and

$$K_n(U,X,\Delta_1,\Delta_2) = \frac{\sigma'}{\sigma}\frac{1}{n}(X'\Delta_1)(X'\Delta_2).$$

The moments conditions (DC) ensure that

$$\frac{1}{\sqrt{n}}\sum_i K_n(U_i,X_i,\Delta_1,\Delta_2) = \frac{1}{\sqrt{n}}\Delta_1'\left(\frac{1}{n}\sum_i \frac{\sigma_i'}{\sigma_i}X_iX_i'\right)\Delta_2 = o_P(1).$$

Using the stochastic equicontinuity arguments in the proof of lemma 2, $\psi_\tau[(U-q(\tau)) - L_n(U,X,\Delta_1,\Delta_2) - 1/\sqrt{n}\Delta_3]$ can by expanded around $\Delta = 0$ to yield,

$$\Delta_3 + \left(\frac{1}{n}\sum_i \frac{1}{\sigma_i}X_i'\right)\Delta_1 + \left(\frac{1}{n}\sum_i \frac{\sigma_i'}{\sigma_i}U_iX_i'\right)\Delta_2 = -\frac{1}{f_U(q(\tau))}\frac{1}{\sqrt{n}}\sum_i \psi_\tau(U_i - q(\tau)) + o_P(1).$$

Consider now Δ_1 and Δ_2. The moment conditions can be written as,

$$\begin{aligned} M_{1_n}(U, X, \Delta) &= \frac{1}{\sqrt{n}} \sum_i C_i(\hat{R}_i / \hat{\sigma}) \\ &= \frac{1}{\sqrt{n}} \sum_i C_i(U_i - L_n(U_i, X_i, \Delta_1, \Delta_2) \\ &= o_P(1), \end{aligned}$$

and

$$\begin{aligned} M_{2_n}(U, X, \Delta) &= \frac{1}{\sqrt{n}} \sum_i C_i(|\hat{R}_i| / \hat{\sigma}) \\ &= \frac{1}{\sqrt{n}} \sum_i 2C_i[U_i - L_n(U_i, X_i, \Delta_1, \Delta_2)] \\ &\quad \times \left[1/2 - I\left\{U_i \le \frac{1}{\sqrt{n}} \frac{1}{\sigma_i} X_i' \Delta_1\right\}\right] \\ &= o_P(1). \end{aligned}$$

As in the proof of lemma 2, the moment conditions ensure the stochastic equicontinuity of $\{M_{2_n}(U, X, \Delta) - E[M_{2_n}(U, X, \Delta)]\}$. Together with the consistency of Δ and the fact that $E[M_{2_n}(U, X, 0)] = 0$, this allows us to write,

$$M_{2_n}(U, X, \Delta) = E[M_{2_n}(U, X, \Delta)] + M_{2_n}(U, X, 0).$$

The linear representation is completed by noting the first-order Taylor series expansion of $E[M_{2_n}(U, X, \Delta)]$ around $\Delta = 0$,

$$E[M_{2_n}(U, X, \Delta)] = E[(1/\sigma)\ \mathrm{sign}(U) C X'] \Delta_1 + E[(\sigma'/\sigma)\, |U| C X'] \Delta_2 + o(1).$$

A final remark about the non-sigularity of G. It suffices to show that

$$\begin{pmatrix} E[(1/\sigma)\, C X'] & E[(\sigma'/\sigma)\, U C X'] \\ E[(1/\sigma)\ \mathrm{sign}(U) C X'] & E[(\sigma'/\sigma)\, |U| C X'] \end{pmatrix}$$

is non-singular which is ensured by (DC5). ■

It is easy to generalize the results of Theorem 5 for multiple quantiles. For $0 < \tau_1 < \tau_2 < \ldots < \tau_m$,

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}-\beta) \\ \sqrt{n}(\hat{\gamma}-\gamma) \\ \sqrt{n}(\hat{q}_1 - q(\tau_1)) \\ \vdots \\ \sqrt{n}(\hat{q}_m - q(\tau_m)) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{G}^{-1}\mathcal{N}(0,\mathcal{H}),$$

where,

$$\mathcal{H} = \begin{pmatrix} E[U^2]E[CC'] & E[UV]E[CC'] & E[C]E[U\Psi(U)'] \\ & E[V^2]E[CC'] & E[C]E[V\Psi(U)'] \\ & & J \end{pmatrix},$$

where

$$\Psi(U) = \left(\frac{\psi_{\tau_1}(U - q(\tau_1))}{f_U(q(\tau_1))} \cdots \frac{\psi_{\tau_m}(U - q(\tau_m))}{f_U(q(\tau_m))} \right)',$$

$J = E[\Psi(U)\Psi(U)']$ is a $m \times m$ matrix with entries $J_{ij} = \frac{\min\{\tau_i,\tau_j\} - \tau_i\tau_j}{f_U(q(\tau_i))f_U(q(\tau_j))}$, and

$$\mathcal{G} = \begin{pmatrix} E[(1/\sigma)\,CX'] & E[(\sigma'/\sigma)\,UCX'] & 0_{k\times m} \\ E[(1/\sigma)\,\text{sign}(U)CX'] & E[(\sigma'/\sigma)\,|U|CX'] & 0_{k\times m} \\ \mathbf{1}_m E[(1/\sigma)\,X'] & \mathbf{1}_m E[(\sigma'/\sigma)\,UX'] & I_{m\times m} \end{pmatrix},$$

where $\mathbf{1}_m$ is a vector of 1s of dimension m and $I_{m\times m}$ is an identity matrix of order m.

REFERENCES

Abadie, A., Angrist, J. and Imbens, G. (2002). "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Andrews, D.W.K. (1994). "Empirical Process Methods in Econometrics," in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, 2248-2294, Amsterdam: Elsevier.

Bassett Jr., G.S. and Koenker, R. (2018). "A Quantile Regression Memoir," in Koenker, R., Chernozhukov, V., He, X., and Peng, L. (eds.), *Handbook of Quantile Regression*, Ch. 1, 3–5, Boca Raton (FL): Chapman-Hall/CRC.

Buchinsky, M. (1994). "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62, 405–458.

Cade, B.S. and Noon, B.R. (2003). "A Gentle Introduction to Quantile Regression for Ecologists," *Frontiers in Ecology and the Environment*, 1, 412–420.

Canay, I.A. (2011). "A Simple Approach to Quantile Regression for Panel Data," *The Econometrics Journal*, 14, 368–386.

Chamberlain, G. (1994). "Quantile Regression, Censoring and the Structure of Wages," in Sims, C.A. (ed.), *Advances in Econometrics*, 171-209, Cambridge: Cambridge University Press.

Chesher, A. (2017). "Understanding the Effect of Measurement Error on Quantile Regressions," *Journal of Econometrics*, 200, 223–237.

Chen, L.-Y. and Lee, S. (2017). *Exact Computation of GMM Estimators for Instrumental Variable Quantile Regression Models*, available at: `arXiv:1703.09382v1`.

Chernozhukov V.; Fernández-Val, I. and Galichon, A. (2010). "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V. and Hansen, C. (2005). "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V. and Hansen, C. (2006). "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, 132, 491–525.

Chernozhukov, V. and Hansen, C. (2008). "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, 142, 379–398.

Chernozhukov, V. and Hong, H. (2003). "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346.

Cochran, W.G. (1952). "The χ^2 Test of Goodness of Fit," *Annals of Mathematical Statistics*, 23, 315–45.

Dhaene, G. and Jochmans, K. (2015). “Split-panel Jackknife Estimation of Fixed-effect Models,” *The Review of Economic Studies*, 82, 991–1030.

Engle, R.F. (2001). “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics,” *Journal of Economic Perspectives*, 15, 157–168.

Fernández-Val, I. and Weidner, M. (2016). “Individual and Time Effects in Nonlinear Panel Models with Large N, T,” *Journal of Econometrics*, 192, 291-312.

Galvão, A.F. (2011). “Quantile Regression for Dynamic Panel Data with Fixed Effects,” *Journal of Econometrics*, 164, 142–157.

Galvão, A.F. and Kato, K. (2016). “Smoothed Quantile Regression for Panel Data,” *Journal of Econometrics* 193, 92-112.

Galvão, A.F. and Kato, K. (2018). “Quantile Regression Methods for Longitudinal Data,” in Koenker, R., Chernozhukov, V., He, X., and Peng, L. (eds.), *Handbook of Quantile Regression*, Ch. 19, 363-380, Boca Raton (FL): Chapman-Hall/CRC.

Galvão, A.F. and Wang, L. (2015). “Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data,” *Journal of Multivariate Analysis*, 133, 1-26.

Glejser, H. (1969). “A New Test for Heteroskedasticity,” *Journal of the American Statistical Association*, 64, 315–323.

Gutenbrunner, C. and Jurečková, J. (1992). “Regression Rank Scores and Regression Quantiles,” *Annals of Statistics*, 20, 305–330.

Hahn, J. and Newey, W. (2004). “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models,” *Econometrica*, 72, 1295–1319.

Hansen, L.P. (1982). “Large Sample Properties of Generalized Methods of Moments Estimators,” *Econometrica*, 50, 1029–1054.

Harvey, A. (1976). “Estimating Regression Models with Multiplicative Heteroscedasticity,” *Econometrica*, 44, 461–465.

He, X. (1997). “Quantile Curves Without Crossing,” *The American Statistician*, 51, 186–192.

Im, K.S. (2000). "Robustifying the Glejser test of Heteroskedasticity," *Journal of Econometrics*, 97, 179–188.

Kaplan, D.M. and Sun, Y. (2017). "Smoothed Estimating Equations for Instrumental Variables Quantile Regression," *Econometric Theory*, 33, 105–157.

Kato, K., Galvão, A.F. and Montes-Rojas, G. (2012). "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

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

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

Koenker, R. and Bassett, G. (1982). "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.

Koenker, R. and Hallock, K.F. (2001). "Quantile Regression," *Journal of Economic Perspectives*, 15, 143–156.

Koenker, R. and Zhao, Q. (1994). "L-Estimatton for Linear Heteroscedastic Models," *Journal of Nonparametric Statistics*, 3, 223–235.

Koenker, R. and Zhao, Q. (1996). "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813.

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

Lancaster, T. (2000). "The Incidental Parameters Problem Since 1948," *Journal of Econometrics*, 95, 391–414.

Leamer, E.E. (2010). "Tantalus on the Road to Asymptotia," *Journal of Economic Perspectives*, 24, 31–46.

Machado, J.A.F. and Santos Silva, J.M.C. (2000), "Glejser's Test Revisited," *Journal of Econometrics*, 97, 189–202.

Newey, W.K. (1985). "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047–1070.

Newey, W.K. and McFadden, D. (1994). "Large Sample Estimation and Hypothesis Testing," in Engle, R.F. and McFadden, D. (eds.), *Handbook of Econometrics*, Vol. 4, Ch. 36, 2111–2245, Amsterdam: Elsevier.

Neyman, J. and Scott, E. (1948). "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

Persson, T. and Tabellini, G.E. (2003). *The Economic Effects of Constitutions*, Cambridge(MA): The MIT Press.

Powell, D. (2016). *Quantile Regression with Nonadditive Fixed Effects*, available at: `http://works.bepress.com/david_powell/1/`.

Romano, J.P. and Wolf, M. (2017). "Resurrecting Weighted Least Squares," *Journal of Econometrics*, 197, 1–19.

StataCorp. (2017). *Stata Statistical Software: Release 15*, College Station: TX: StataCorp LLC.

Wooldridge, J.M. (1999). "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, 90, 77–97.

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge(MA): The MIT Press.

Wüthrich, K. (2017). "Semiparametric Estimation of Quantile Treatment Effects with Endogeneity," Discussion Papers, Universität Bern, Department of Economics, No. 15-09.

Xu, G. and Burer, S. (2017). "A Branch-and-Bound Algorithm for Instrumental Variable Quantile Regression," *Mathematical Programming Computation*, 9, 1–27.

Zhao, Q. (2000). "Restricted Regression Quantiles," *Journal of Multivariate Analysis*, 72, 78–99.