Two-stage quantile regression when the first stage is based on quantile regression

TAE-HWAN KIM[*,†] AND CHRISTOPHE MULLER[‡]

School of Economics, University of Nottingham, Nottingham NG7 2RD, U.K.

†*Department of Economics, Yonsei University, Seoul 120-749, Korea*
E-mail: Tae-Hwan.Kim@nottingham.ac.uk

‡*Departamento de Fundamentos del Análisis Económico, Universidad de Alicante, Campus de San vicente, 03080 Alicante, Spain*
E-mail: cmuller@merlin.fae.us.es

Summary   We present the asymptotic properties of double-stage quantile regression estimators with random regressors, where the first stage is based on quantile regressions with the same quantile as in the second stage, which ensures robustness of the estimation procedure. We derive invariance properties with respect to the reformulation of the dependent variable. We propose a consistent estimator of the variance–covariance matrix of the new estimator. Finally, we investigate finite sample properties of this estimator by using Monte Carlo simulations.

Keywords: *Two-stage estimation, Quantile regression, Endogeneity.*

1. INTRODUCTION

Quantile regressions are often used not only for wage and living standard analyses[1] but also for studies of firm data and financial data. Quantile regressions produce robust estimates, particularly for misspecification errors related to non-normality and to the presence of outliers. They also help the researcher to focus his/her study on specific parts of the conditional distribution.

The researchers often study relations in which some right-hand-side variables are endogenous. For example, socioeconomic variables, such as the education of the individual, appearing in wage equations may be endogenous. Other sources of endogeneity such as measurement errors are common.

When there are endogenous variables, the estimator of the parameter of interest is generally inconsistent. A well-known solution is the two-stage least squares (2SLS) estimation method in which one replaces the endogenous explanatory variables with their predictions from ancillary equations based on other exogenous variables. However, if researchers are interested in a specific part of the distribution of the structural variable, other than the mean, or if they want robust estimates, the 2SLS estimation method is not appropriate. Amemiya (1982) and Powell (1983)

---

[1]See Buchinsky (1995) and Machado and Mata (2001).

dealt with the case of the double-stage least-absolute deviations (DSLAD) with fixed regressors, which allow researchers to focus on the median of the distribution of interest. In this paper, we extend the works of Amemiya and Powell by using quantile regressions and random exogenous variables. We use the same quantile estimation in two steps and the resulting estimator of the structural parameter is termed 'double-stage quantile regression (DSQR) estimator'.

Other researchers have treated some endogeneity problems in quantile regressions. Chen and Portnoy (1996) study two-stage quantile regressions with symmetric error terms where the first-stage estimators are trimmed least-squares estimators and LAD estimators. Other authors do not rely on the simple two-stage parametric approach favoured by many empirical economists. Kemp (1999) and Sakata (2001) studied least absolute error difference (LAED) estimators for estimating a single equation from a simultaneous equation model.[2] Abadie *et al.* (2002) design a quantile treatment effects estimator, which is the solution to a convex programming problem with first-step non-parametric estimation of a nuisance function.[3] MaCurdy and Timmins (2000) propose an estimator for ARMA models adapted to the quantile regression framework.[4]

We do not follow these various approaches in this paper and we rather focus on two-stage estimation procedures familiar to empirical economists. Some empirical economists (Arias *et al.* 2001; Garcia *et al.* 2001). adopt a direct approach with a first stage of LS estimators. However, this approach may be delicate for the general type of problem that we consider since using LS estimation in the first step combined with Amemiya's reformulation of the dependent variable may produce an asymptotic bias (as will be shown in Appendix A). It can also destroy the robustness properties of the quantile regressions. We will focus on the case when the first stage is composed of quantile regressions for which the quantile is the same as for the second stage.

This paper is organised as follows. Section 2 discusses the model and the assumptions. At this occasion, we derive invariance properties with respect to the reformulation of the dependent variable. In Section 3, we prove the asymptotic normality of the DSQR and discuss the estimation of the asymptotic variance–covariance matrix. We emphasise that the same quantile should be used to estimate the reduced and structural forms in order to avoid asymptotic biases for all values of q. We present simulation results in Section 4. Finally, Section 5 concludes. The proofs are presented in Appendix B.

2. THE MODEL

We are interested in the structural parameter, $\alpha_0 = (\gamma_0', \beta_0')'$, in an equation in the following matrix form for a sample of T observations:

$$y = Y\gamma_0 + X_1\beta_0 + u, \tag{1}$$

where $[y, Y]$ is a $T \times (G + 1)$ matrix of endogenous variables, X_1 is $T \times K_1$ matrix of exogenous variables and u is a $T \times 1$ vector. The matrix X_2 contains $K_2(=K - K_1)$ exogenous variables absent from (1). In this situation the endogeneity of Y in (1) may cause that $Q_\theta(u|Y) \neq Q_\theta(u)$,

---

[2]These estimators are not based on first-stage estimates of the reduced-form equations for the right-hand-side endogenous variables. Instead, they use encompassed optimisation procedures, in which explicit concentration of formulae is impossible. The LAED optimisation is not a linear programming problem and a grid search must be used for the concentration.

[3]They deal with the case of binary treatment related to sample selectivity by modifying the typical objective function of the quantile regression problem with non-parametrically estimated weights.

[4]It is based on the smoothing of conditional quantile conditions that are incorporated in a GMM framework.

where $Q_\theta(\cdot)$ is the quantile function of order θ, and $Q_\theta(\cdot|Y)$ is the quantile function of order θ conditional on Y. This non-equality can be used as a definition of endogeneity for quantile regressions. Moreover, we assume that Y has the following reduced-form representation:

$$Y = X\Pi_0 + V, \tag{2}$$

where $X = [X_1, X_2]$ is a $T \times K$ matrix, Π_0 is a $K \times G$ matrix of unknown parameters and V is a $T \times G$ matrix of unknown error terms. Then, the reduced-form representation of y is

$$y = X\pi_0 + v, \tag{3}$$

where $\pi_0 = [\Pi_0, \binom{1_{K_1}}{0}]\alpha_0 = H(\Pi_0)\alpha_0$ and $v = u + V\gamma_0$.

Define $\rho_\theta : R \to R^+$ for given $\theta \in (0, 1)$ as $\rho_\theta(z) = z\psi_\theta(z)$, where $\psi_\theta(z) = \theta - 1_{[z \le 0]}$ and $1_{[.]}$ is the Kronecker index. As a natural extension of Amemiya (1982) and Powell (1983), we define the double-stage quantile regression estimator (DSQR(θ, q)) $\hat{\alpha} = (\hat{\gamma}', \hat{\beta}')'$ of α_0 as a solution to the following minimisation programme:

$$\min_\alpha S_T(\alpha, \hat{\pi}, \hat{\Pi}, q, \theta) = \sum_{t=1}^{T} \rho_\theta(qy_t + (1-q)x_t'\hat{\pi} - x_t'H(\hat{\Pi})\alpha), \tag{4}$$

where y_t and x_t' are the tth elements in y and X, respectively, q is a positive constant chosen in advance by the researcher, and $\hat{\pi}$ and $\hat{\Pi}_j$ (the jth column of $\hat{\Pi}$) are the first-stage estimators obtained by

$$\min_\pi \sum_{t=1}^{T} \rho_\theta(y_t - x_t'\pi) \text{ and } \min_{\Pi_j} \sum_{t=1}^{T} \rho_\theta(Y_{jt} - x_t'\Pi_j), (j = 1, \ldots, G), \tag{5}$$

where π and Π_j are $K \times 1$ vectors and Y_{jt} is the (j, t)th element of Y. The reformulation of the dependent variable as $qy_t + (1-q)x_t'\hat{\pi}$ has been introduced by Amemiya as a generalisation of a property of 2SLS, and an attempt to improve efficiency. Although the ability of choosing the value of q should yield estimators depending on this value, we show in the next proposition that there exist cases where the DSQR(θ, q) is invariant with respect to the value of q.

Proposition 1 *Let $\delta(\theta, y, X)$ denote the quantile regression estimator associated with quantile θ, dependent variable y and matrix of right-hand-side variables X. Moreover, let $\tilde{\alpha} = \delta(\theta, y, XH(\hat{\Pi}))$. Then, we have*

(i) $\delta(\theta, qy + (1-q)XH(\hat{\Pi})\tilde{\alpha}, XH(\hat{\Pi})) = \tilde{\alpha}$.
(ii) *If $K_2 = G$ and $H(\hat{\Pi})$ is of full column rank, then the DSQR(θ, q) is given by $\hat{\alpha} = H(\hat{\Pi})^{-1}\hat{\pi}$.*

The first result (i) is an invariance property of $\hat{\alpha}$ that is also verified by least square estimators. Although it does not correspond to the composite dependent variable that we will consider later on, this result implies that iterating the estimation by changing the dependent variable in that way is useless, since the initial DSQR$(\theta, 1)$ is obtained. The second property (ii) isolates the case of exact identification, where the distribution of the estimator, and in fact the estimator itself, do not depend on q. The argument and the direct calculus here are analogous to that used for indirect least-squares (ILS) estimator. Result (ii) shows how easily the DSQR(θ, q) can be obtained when the exact identification condition is satisfied. In that case, no numerical minimisation of (4) is necessary and this is precisely in the same way in which ILS is obtained. Our theoretical results for the DSQR(θ, q) developed in the subsequent sections are general enough to include not only the exactly identified case ($K_2 = G$), but also the over-identified case ($K_2 > G$).

4   *Tae-Hwan Kim and Christophe Muller*

The following conditions will be useful to obtain the asymptotic representation of the DSQR(θ, q), which we discuss in Appendix A, and the asymptotic normality of the DSQR(θ, q).

Assumption 1 *The sequence* $\{(v_t, V_t, x_t)\}$ *is independent and identically distributed (i.i.d.) where* u_t *and* V_t *are the t*th *elements in u and V, respectively.*

Assumption 2 (i) $E(||x_t||^3) < \infty$.

(ii) $H(\Pi_0)$ *is of full column rank.*

(iii) *The conditional densities* $f(\cdot|x)$ *and* $g_j(\cdot|x)$, *respectively, for* v_t *and* V_{jt}, *are Lipschitz continuous for all x. Moreover,* $Q_0 = E\{f(0|x_t)x_tx_t'\}$ *and* $Q_j = E\{g_j(0|x_t)x_tx_t'\}$ *are finite and positive definite.*

(iv) $E\{\psi_\theta(v_t)|x_t\} = 0$ *and* $E\{\psi_\theta(V_{jt})|x_t\} = 0$ *($j = 1, \ldots, G$).*

Assumption 2(i) is needed to generalise the stochastic equicontinuity result in Powell (1983). Assumptions 2(ii) and 2(iii) are standard in the literature. Assumption 2(iv) is a generalisation of Powell's assumption and states that zero is the quantile of order θ of v_t and V_{jt} conditionally on x_t. Assumption 2(iv) normalises the intercept on the θth quantile of the distributions of v_t and V_{jt}. The occurrence of a bias, when different quantiles are used for the two stages or when least-square estimators are used for the first stage, is discussed in Appendix A. For $q \neq 1$, distinct methods for first and second stages generally imply incompatible restrictions on error terms. We are now ready to study the asymptotic normality of the DSQR(θ, q).

3. ASYMPTOTIC NORMALITY AND COVARIANCE MATRIX

The asymptotic normality of the DSQR(θ, q) is based on its asymptotic representation derived in Appendix A. It is easy to see from this representation in (15) that the DSQR(θ, q) is asymptotically robust because its influence function is a linear combination of bounded functions. The robustness would be lost if non-robust first-stage estimators were used.

Proposition 2 *Under Assumptions 1 and 2,* $T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, D\Omega D')$, *where* $D = Q_{zz}^{-1}H(\Pi_0)'[I, -Q_0Q_1^{-1}\gamma_{01}, \ldots, -Q_0Q_G^{-1}\gamma_{0G}]$, $Q_{zz} = H(\Pi_0)'Q_0H(\Pi_0)$ *and* $\Omega = E(\Sigma \otimes x_tx_t')$, *where* Σ *is the matrix of general term* $\psi_\theta(W_{it})\psi_\theta(W_{jt})$ *with* $W_{1t} = v_t$, $W_{it} = V_{i-1,t}$ *for* $2 \leq i \leq G + 1$.

Note that although the definition of $\hat{\alpha}$ depends on q, its asymptotic law does not. In that case the first-stage estimator $\hat{\pi}$ intervenes in the calculation of $\hat{\alpha}$, but not in its asymptotic distribution. However, $\hat{\pi}$ can still be used for a consistent estimator of the asymptotic covariance matrix. In order to conduct inference based on $\hat{\alpha}$, it is necessary to have consistent estimators of D and Ω. We now develop such estimators by using the plug-in principle.

Assumption 3 (i) *There exist positive constants* f_0 *and* g_{0j} *such that* $f(\cdot|x) \leq f_0$ *and* $g_j(\cdot|x) \leq g_{0j}$ *for all x.*

(ii) *There exists a stochastic sequence* $\{\hat{c}_{jT}\}$ *and a non-stochastic sequence* $\{c_{jT}\}$ *for* $j = 0, 1, \ldots, G$ *such that we have* $\hat{c}_{jT}/c_{jT} \xrightarrow{p} 1$, $c_{jT} = o_p(1)$ *and* $c_{jT}^{-1} = o_p(T^{1/2})$.

We defined our covariance estimators as follows:

$$\hat{D} = \hat{Q}_{zz}^{-1}H(\hat{\Pi})'\left[I, -\hat{Q}_0\hat{Q}_1^{-1}\hat{\gamma}_1, \ldots, -\hat{Q}_0\hat{Q}_G^{-1}\hat{\gamma}_G\right],$$

where $\hat{Q}_{zz} = H(\hat{\Pi})' \hat{Q}_0 H(\hat{\Pi})$, $\hat{Q}_0 = (2\hat{c}_{0T}T)^{-1} \sum_{t=1}^{T} 1_{[-\hat{c}_{0T} \leq \hat{v}_t \leq \hat{c}_{0T}]} x_t x_t'$, $\hat{Q}_j = (2\hat{c}_{jT}T)^{-1} \sum_{t=1}^{T}$ $1_{[-\hat{c}_{jT} \leq \hat{V}_{jt} \leq \hat{c}_{jT}]} x_t x_t'$, $\hat{v}_t = y_t - x_t'\hat{\pi}$ and $\hat{V}_{jt} = Y_{jt} - x_t'\hat{\Pi}(j = 1, \ldots, G)$;

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \hat{\Sigma}_t \otimes x_t x_t',$$

where $\hat{\Sigma}_t = \hat{W} \otimes x_t x_t'$ and $\hat{W}$ is the matrix of general term $\psi_\theta(\hat{W}_{it})\psi_\theta(\hat{W}_{jt})$ with $\hat{W}_{1t} = \hat{v}_t$, $\hat{W}_{it} = \hat{V}_{i-1,t}$ for $2 \leq i \leq G + 1$. Then, our proposed covariance matrix estimator is $\hat{D}\hat{\Omega}\hat{D}'$.

Proposition 3 *Under Assumptions 1–3, $\hat{D} \xrightarrow{p} D$ and $\hat{\Omega} \xrightarrow{p} \Omega$.*

To the best of our knowledge that is the first time that a proof of convergence of the variance–covariance matrix of the two-stage quantile estimators is given and even with non-random regressors. We now complete these asymptotic results with small sample simulations.

4. MONTE CARLO SIMULATIONS

In this section, we conduct simulation experiments to investigate the finite-sample properties of the DSQR(θ, q),[5] notably in terms of the endogeneity and robustness problems.

The data-generating process used in the simulations is described in Appendix C, which shows that the equation of interest is over-identified and the parameter values are $\gamma_0 = 0.5$ and $\beta_0' = (1, 0.2)$. We generate the error terms by using three alternative distributions: the standard normal N(0,1), the Student-*t* with 3 degrees of freedom $t(3)$ and the Lognormal LN(0,1). The exogenous variables x_t are also drawn in a normal distribution at each replication. The number of replications is 1,000. For each replication, we estimate the parameter values γ_0 and β_0 and the deviation of the estimates from the true values. Then, we compute the sample mean and sample standard deviation (when useful, the sample median and the sample interquartile range) over the 1,000 replications.

The performance of the one-stage quantile regression estimator for the different distributions is displayed in Table 1 for the N(0,1) case. The results for the other distributions are qualitatively similar and are not reported. This estimator exhibits a systematic bias in finite samples, which does not disappear as the sample size increases from 50 to 300.

The results for the DSQR(θ, q) with N(0,1) are provided in Tables 2 and 3. We select two illustrative values (0.5 and 1) for q. As shown in Proposition 1, when the system is exactly identified, this dependence on q does not exist, which is confirmed by our simulations of this case (not shown). Whereas the DSQR(θ, q) does not depend on q asymptotically, it does in finite samples, but as we increase the sample size to 300, the results for different q's become quantitatively similar. The means of the DSQR(θ, q) estimates, ($\hat{\gamma} - \gamma_0$, $\hat{\beta} - \beta_0$), are much closer to zero than the means of the one-step quantile estimator over all values of θ, although the corresponding standard deviations are generally greater.

For other distributions too, the changes in the value of q do not substantially modify the results. Hence, for these two distributions, $t(3)$ and LN(0,1), we report only the case $q = 1$. The

Q1

---

[5]See also Ribeiro (1998) for small sample simulations of two-stage LAD estimation with a first step of LS or LAD estimators.

Table 1. Simulation means and standard deviations of one-step quantile estimator: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\tilde{\gamma}$	Mean	−0.41	−0.40	−0.40	−0.40	−0.41
		Std	0.28	0.18	0.16	0.17	0.28
	$\tilde{\beta}_0$	Mean	0.90	0.96	1.05	1.14	1.26
		Std	1.23	0.62	0.45	0.39	0.42
	$\tilde{\beta}_1$	Mean	0.09	0.09	0.09	0.09	0.08
		Std	0.34	0.21	0.19	0.21	0.33
$T = 300$	$\tilde{\gamma}$	Mean	−0.41	−0.41	−0.41	−0.41	−0.41
		Std	0.11	0.07	0.06	0.07	0.11
	$\tilde{\beta}_0$	Mean	0.84	0.97	1.06	1.17	1.30
		Std	0.47	0.23	0.18	0.16	0.17
	$\tilde{\beta}_1$	Mean	0.09	0.09	0.09	0.09	0.09
		Std	0.13	0.09	0.08	0.09	0.13

Table 2. Simulation means and standard deviations of $DSQR(\theta, q = 0.5) : N(0,1)$.

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\tilde{\gamma}$	Mean	−0.01	−0.01	−0.01	−0.00	−0.01
		Std	0.71	0.45	0.33	0.37	0.72
	$\tilde{\beta}_0$	Mean	−0.04	−0.06	−0.01	−0.00	−0.10
		Std	1.96	1.17	0.87	0.99	1.85
	$\tilde{\beta}_1$	Mean	0.01	0.00	0.00	0.00	0.01
		Std	0.45	0.25	0.22	0.25	0.43
$T = 300$	$\tilde{\gamma}$	Mean	0.01	0.00	0.01	0.00	0.00
		Std	0.21	0.13	0.12	0.13	0.20
	$\tilde{\beta}_0$	Mean	−0.03	−0.01	−0.02	−0.01	−0.01
		Std	0.58	0.35	0.33	0.36	0.54
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.15	0.09	0.09	0.10	0.14

results with $t(3)$, which are given in Table 4 show similar features. However, the fatter tails of the errors entail accuracy losses for both one-stage and two-stage estimators. The results with lognormal error terms, shown in Table 5, differ in that both estimators are severely biased for large quantiles (for $\theta = 0.95$). The bias of the DSQR(θ, q) diminishes when the sample size rises to $T = 300$. In a simulation available upon request, the bias disappears for a sufficiently large number of observations, as opposed to the case of the one-stage estimator. Also, the performance of the DSQR(θ, q) for the lognormal case is the best for small quantiles ($\theta = 0.05$), in contrast with the usual better performance of the DSQR(θ, q) for quantiles around $\theta = 0.5$ and symmetric distributions. The formula of the diagonal term of the covariance matrix $D\Omega D'$ suggests us a conjecture for the occurrence of such effect in large samples, which may extend to small samples in some cases. Indeed, because of the asymmetric shape of the lognormal distribution, $f(0|x_t)$ and

Table 3. Simulation means and standard deviations of $DSQR(\theta, q = 1) : N(0,1)$.

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\tilde{\gamma}$	Mean	−0.01	0.01	0.00	0.00	−0.01
		Std	0.75	0.45	0.33	0.39	0.76
	$\tilde{\beta}_0$	Mean	0.00	0.04	0.00	−0.01	0.03
		Std	2.08	−1.17	0.88	1.04	1.94
	$\tilde{\beta}_1$	Mean	0.01	0.00	0.00	0.00	0.01
		Std	0.47	0.26	0.22	0.25	0.44
$T = 300$	$\tilde{\gamma}$	Mean	0.01	0.01	0.01	0.00	0.00
		Std	0.22	0.13	0.12	0.13	0.20
	$\tilde{\beta}_0$	Mean	−0.02	−0.01	−0.02	−0.01	0.00
		Std	0.59	0.36	0.34	0.36	0.55
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.15	0.09	0.09	0.10	0.15

Table 4. Simulation means and standard deviations of $DSQR(\theta, q = 1) : t(3)$.

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\tilde{\gamma}$	Mean	−0.31	−0.01	0.00	−0.05	−0.26
		Std	1.49	0.49	0.38	0.57	1.29
	$\tilde{\beta}_0$	Mean	0.61	0.01	0.01	0.16	0.92
		Std	4.16	1.32	1.01	1.54	3.41
	$\tilde{\beta}_1$	Mean	0.09	−0.01	0.00	0.01	0.05
		Std	1.11	0.33	0.26	0.35	1.04
$T = 300$	$\tilde{\gamma}$	Mean	−0.03	0.00	−0.01	0.00	−0.05
		Std	0.57	0.17	0.13	0.17	0.86
	$\tilde{\beta}_0$	Mean	0.06	0.00	0.03	0.01	0.18
		Std	1.48	0.45	0.36	0.45	2.74
	$\tilde{\beta}_1$	Mean	0.01	−0.01	0.00	−0.01	0.01
		Std	0.35	0.12	0.10	0.12	0.38

$g_j(0|x_t)$ for large values of θ when the θth quantile is zero can be very close to zero. This inflates the roles of some terms in the diagonal terms of $D\Omega D'$, which may lead to a large variance for the DSQR(θ, q) and cause the large finite-sample bias for $\theta = 0.95$ and LN(0,1).

For all the types of error terms, the one-stage estimator is severely biased. In contrast, the DSQR(θ, q) has good finite sample properties, although a too small sample size or too extreme quantiles ($\theta = 0.05, 0.95$) may degrade its performance. Sample medians and sample interquartile ranges have been calculated to supplement sample means and sample standard deviations, respectively. We have not found any significant difference between the robust measures and the usual measures, except in the case where the error terms are drawn from LN(0,1) and with large quantiles ($\theta = 0.95$). In that case only, the robust measures are reported in the brackets, next to their corresponding usual measures in Table 4. The dispersion of the sampling distribution of the deviations is smaller when robust measures are used.

Table 5. Simulation means and standard deviations of $DSQR(\theta, q = 1)$: LN(0,1).

			0.05	0.25	0.50	0.75	0.95
		θ					
$T = 50$	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.06	0.47(0.45)
		Std	0.07	0.09	0.44	0.41	0.55(0.35)
	$\tilde{\beta}_0$	Mean	−0.06	−0.05	−0.02	0.05	0.59(0.67)
		Std	0.05	0.08	0.15	0.33	1.11(0.71)
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	−0.01	−0.12(−0.13)
		Std	0.05	0.07	0.12	0.23	0.56(0.51)
$T = 300$	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.24(0.26)
		Std	0.02	0.03	0.06	0.13	0.69(0.33)
	$\tilde{\beta}_0$	Mean	−0.08	−0.06	−0.03	0.02	0.40(0.44)
		Std	0.02	0.03	0.05	0.13	0.86(0.58)
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	−0.05(−0.06)
		Std	0.01	0.02	0.04	0.10	0.43(0.34)

Table 6. Simulation means and standard deviations of $DSQR(\theta, q = 1)$ and $DSLS$ with a single outlier: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95	*DSLS*
$T = 50$	$\tilde{\gamma}$	Mean	−0.01	0.01	0.01	0.01	0.08	0.23
		Std	0.77	0.45	0.34	0.40	1.20	1.57
	$\tilde{\beta}_0$	Mean	0.00	−0.02	0.01	0.02	0.27	0.26
		Std	2.13	1.18	0.91	1.08	3.07	4.02
	$\tilde{\beta}_1$	Mean	0.01	0.00	0.00	0.00	−0.01	0.01
		Std	0.47	0.26	0.23	0.27	1.10	1.17
$T = 300$	$\tilde{\gamma}$	Mean	0.01	0.01	0.01	0.00	0.01	0.04
		Std	0.22	0.13	0.12	0.13	0.21	0.25
	$\tilde{\beta}_0$	Mean	−0.02	−0.01	−0.01	−0.01	0.01	0.05
		Std	0.59	0.36	0.34	0.37	0.57	0.65
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.01	0.01
		Std	0.15	0.10	0.09	0.10	0.15	0.19

One of the justifications for using the DSQR(θ, q) is that it is resistant to outliers. To confirm this property, we conduct a separate Monte Carlo experiment in which we compare the DSQR(θ, q) with the 2SLS when one outlier occurs for the dependent variable y_t. Following Cowell and Flachaire (2002), the outlier is the randomly selected observation of y_t and multiplied by 15. The results in the normal error case are reported in Table 6 for the 2SLS, which is invariant to the value of q, and for the DSQR(θ, 1). Other distributions and the other values of q deliver similar results. The results show that the DSQR(θ, q) is much more robust to outliers than the 2SLS, which is still more obvious when medians and interquartiles indicators are used.

5. CONCLUSION

We study in this paper the asymptotic properties, the invariance, the robustness and the finite sample properties of double-stage quantile regression estimators with first-stage quantile regressions defined on the same quantile as the second stage. Our results permit valid inferences in models estimated using quantile regressions with random regressors, in which the possible endogeneity of some explanatory variables is treated via preliminary predictive quantile regressions.

ACKNOWLEDGEMENTS

REFERENCES

Abadie, A., J. Angrist and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*, 91–117.

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica 50*, 689–711.

Andrews, D. W. K. (1994). Empirical process methods in econometrics. *Handbook of Econometrics*, Volume IV, eds. R. F. Engle and D. L. McFadden, New York: North-Holland, pp. 2247–94.

Arias, O., K. F. Hallock and W. Sosa-Escudero (2001). Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data. *Empirical Economics 26*, 7–40.

Buchinsky, M. (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics 65*, 109–54.

Chen, L.-A. and S. Portnoy (1996). Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Communications Statistics-Theory Methods 25*, 1005–32.

Cowell, F. and E. Flachaire (2000). Sensitivity of inequality measures to extreme values. Mimeo London School of Economics.   **Q2**

Garcia, J., P. J. Hernandez and A. Lopez-Nicolàs (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics 26*, 149–67.

Kemp, G. (1999). Least absolute error difference estimation of a single equation from a simultaneous equations system. Mimeo, University of Essex.

Kim, T. and C. Muller (2000). Two-stage quantile regression. Nottingham University Working Paper.

Kim, T. and C. Muller (2003). Two-stage quantile regression when the first stage is based on quantile regression. Nottingham University Working Paper.

Koenker, R. and Q. Zhao (1994). L-estimation for linear heteroscedastic models. *Journal of Non-parametric Statistics 3*, 223–235.

Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory 12*, 793–813.

Machado, J. A. F. and J. Mata (2001). Earning functions in Portugal 1982–1994: Evidence from quantile regressions. *Empirical Economics 26*, 115–34.

MaCurdy, T. and C. Timmins (2000). Bounding the influence of attrition on intertemporal wage variation in the NLSY. Mimeo Stanford University.

Pollard, D. (1991). Asymptotics for least absolute deviations regression estimators. *Econometric Theory 7*, 186–199.

Powell, J. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica 51*, 1569–75.

Ribeiro, E. (1998). Small sample evidence of quantile regression for structural models: Estimation and testing. *Revista de Econometria 18*, 215–44.

Sakata, S. (2001). Instrumental variable estimation based on mean absolute deviation. Mimeo University of Michigan.

A. APPENDIX. ASYMPTOTIC REPRESENTATION

We define an empirical process $M_T^*(\zeta)$ by

$$M_T^*(\zeta) = T^{-1/2} \sum_{t=1}^{T} x_t \psi_\theta(qv_t - x_t'\zeta) = T^{-1/2} \sum_{t=1}^{T} m^*(w_t, \zeta),$$

where ζ is a $K \times 1$ vector, $w_t = (v_t, x_t')'$, $m^*(w_t, \zeta) = x_t \psi_\theta(qv_t - x_t'\zeta)$ and $\zeta = T^{-1/2}\delta \in R^K$ with the norm $||.||$. Let $V_T^*(\zeta) = T^{-1/2} \sum_{t=1}^{T} [m^*(w_t, \zeta) - E\{m^*(w_t, \zeta)\}] = M_T^*(\zeta) - E(M_T^*(\zeta))$. Since the function ψ_θ is of bounded variations, Assumptions 1 and 2 are sufficient to apply Theorems 1–3 in Andrews (1994), which leads to:

$$\sup_{||\zeta_1 - \zeta_2|| \leq L^*} \left|\left| V_T^*(\zeta_1) - V_T^*(\zeta_2) \right|\right| = o_p(1) \tag{A.1}$$

for any finite and positive scalar L^*. We outline the proof below. In order to apply Theorem 1 in Andrews (1994), the following two conditions must be verified:

1. $m^*(w_t, \zeta)$ satisfies Pollard's entropy condition with some envelop $\bar{M}(w_t)$;
2. For some $\delta > 2$, $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E[\{\bar{M}(w_t)\}^\delta] < \infty$.

Let $f_1(w_t, \zeta) = x_t$ and $f_2(w_t, \zeta) = \psi_\theta(qv - x_t'\zeta)$ so that $m^*(w_t, \zeta) = f_1(w_t, \zeta) f_2(w_t, \zeta)$. Since $f_1(\cdot, \zeta)$ and $f_2(\cdot, \zeta)$ are Type I functions with envelopes $||x_t||$ and 1, respectively (see Andrews, 1994, for definitions), $m^*(w_t, \zeta)$ satisfies Pollard's entropy condition with envelope $\max(1, ||x_t||)$ because it is a product of two Type I functions (See Theorems 2 and 3 in Andrews, 1994). Hence, (i) is verified. The second condition is now given by $\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E[\{\bar{M}(w_t)\}^\delta] = E[\{\max(1, ||x_t||)\}^\delta]$ which is bounded by Assumption 2(i). Hence, the result in (A.1) now follows.

For any ζ_1, ζ_2 in R^K and $L^*(>0)$ in R, we can find Δ_1, Δ_2 in R^K and $L(>0)$ in R such that $\zeta_1 = T^{-1/2}\Delta_1$, $\zeta_2 = T^{-1/2}\Delta_2$ and $L^* = T^{-1/2}L$. With these notations, we define the following functions: $V_T(\Delta) = M_T(\Delta) - E(M_T(\Delta))$ where $M_T(\Delta) = T^{-1/2} \sum_{t=1}^{T} m(w_t, \Delta)$ and $m(w_t, \Delta) = x_t \psi_\theta(qv_t - T^{-1/2}x_t'\Delta)$. Then, it is easy to see that the result in (A.1) can be written as

$$\sup_{||\Delta_1 - \Delta_2|| \leq L} ||V_T(\Delta_1) - V_T(\Delta_2)|| = o_p(1). \tag{A.2}$$

By setting $\Delta_1 = \Delta$ and $\Delta_2 = 0$ in (A.2), we have

$$\sup_{||\Delta|| \leq L} ||M_T(\Delta) - M_T(0) - \{E(M_T(\Delta)) - E(M_T(0))\}|| = o_p(1). \tag{A.3}$$

Using Assumptions 2(i) and 2(iii), it can be shown that $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$ where $Q_0 = E\{f(0|x_t)x_t x_t'\}$. A detailed proof of this result is provided in Kim and Muller (2003). Therefore, by replacing $E(M_T(\Delta)) - E(M_T(0))$ with its limit $-q^{-1}Q_0\Delta$, we obtain the following result:

$$\sup_{||\Delta|| \leq L} \left|\left| M_T(\Delta) - M_T(0) + q^{-1}Q_0\Delta \right|\right| = o_p(1). \tag{A.4}$$

We now show how the result in (A.4) can be used to introduce the first-stage estimators.

Let $\hat{\Delta}_0 = (q-1)\sqrt{T}(\hat{\pi} - \pi_0) + \sqrt{T}(\hat{\Pi} - \Pi_0)\gamma_0$. Then, $\hat{\Delta}_0 = O_p(1)$ since $\sqrt{T}(\hat{\pi} - \pi_0) = O_p(1)$ and $\sqrt{T}(\hat{\Pi} - \Pi_0) = O_p(1)$, given Assumptions 1 and 2. Hence, from the result in (A.4), we have $M_T(\hat{\Delta}_0) = M_T(0) - q^{-1}Q_0\hat{\Delta}_0 + o_p(1)$, which in turn implies that

$$M_T(\hat{\Delta}_0) = O_p(1) \tag{A.5}$$

since $M_T(0) = O_p(1)$[6] and $q^{-1}Q_0\hat{\Delta}_0 = O_p(1)$. Now let

$$\hat{\Delta}_1(\delta) = H(\hat{\Pi})\delta + \hat{\Delta}_0, \tag{A.6}$$

where $\delta \in R^{G+K_1}$. For some finite $L_1 > 0$, then it is straightforward to show that (A.4) implies that

$$\sup_{||\delta|| \leq L_1} \left|\left| M_T(\hat{\Delta}_1(\delta)) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta) \right|\right| = o_p(1). \tag{A.7}$$

Next, we define $\tilde{M}_T(\delta) = H(\hat{\Pi})'M_T(\hat{\Delta}_1(\delta))$ and $||H(\hat{\Pi})||^2 = tr(H(\hat{\Pi})H(\hat{\Pi})') = O_p(1)$. Using the fact that $\hat{\Pi} - \Pi_0 = o_p(1)$ and the argument between (A.7) and (A.8) in Powell (1983), it is shown that (A.5) and (A.7) together imply that

$$\sup_{||\delta|| \leq L_1} \left|\left| \tilde{M}_T(\delta) - H(\Pi_0)'M_T(\hat{\Delta}_0) + q^{-1}Q_{zz}\delta \right|\right| = o_p(1), \tag{A.8}$$

where $Q_{zz} = H(\Pi_0)'Q_0 H(\Pi_0)$. We now wish to exploit this result with $\hat{\delta} = T^{1/2}(\hat{\alpha} - \alpha)$. For this, we need to show that $\hat{\delta} = O_p(1)$. This is done by using Lemma A.4 in Koenker and Zhao (1996), which can be applied under the following conditions:

  (i) $-\delta'\tilde{M}_T(\lambda\delta) \geq -\delta'\tilde{M}_T(\delta)$ for $\lambda \geq 1$,
  (ii) $||H(\Pi_0)'M_T(\hat{\Delta}_0)|| = O_p(1)$,
  (iii) $\tilde{M}_T(\hat{\delta}) = o_p(1)$, where $\hat{\delta} = T^{1/2}(\hat{\alpha} - \alpha_0)$,
  (iv) Q_{zz} is positive definite.

With these conditions together with $q > 0$, Lemma A.4 in Koenker and Zhao delivers the desired results: $\hat{\delta} = O_p(1)$ and $\hat{\delta} = q Q_{zz}^{-1} H(\Pi_0)'M_T(\hat{\Delta}_0) + o_p(1)$. We first consider condition (i). As noted in the proof of Lemma 3.7 in Koenker and Zhao (1994), the following function $h(\lambda)$ is a convex function:

$$h(\lambda) = \sum_{t=1}^{T} \rho_\theta \left(q v_t - T^{-1/2}x_t'H(\hat{\Pi})\delta\lambda - T^{-1/2}x_t'\hat{\Delta}_0 \right),$$

which implies that the gradient of $h(\lambda)$ given by $-\delta'\tilde{M}_T(\lambda\delta)$ is non-decreasing in λ. Hence, condition (i) is satisfied. Condition (ii) is simply a consequence of (10). To show the validity of (iii), we note that

$$T^{1/2}\tilde{M}_T(\hat{\delta}) = \left[\left. \frac{\partial S_T}{\partial \alpha} \right|_{\alpha=\hat{\alpha}} \right]_-$$

---

[6] Noting that for $q > 0$, $\psi_\theta(qv_t) = \psi_\theta(v_t)$, we have $M_T(0) = T^{-1/2}\sum_{t=1}^{T} x_t\psi_\theta(v_t)$, which, under Assumptions 1, 2(i) and 2(iv), converges in distribution to a normal random variable by the Lindeberg-Levy CLT.

where $[\frac{\partial S_T}{\partial \alpha}|_{\alpha=\hat{\alpha}}]_-$ is $o_p(1)$ because it is the vector of left-hand-side partial derivatives of the objective function S_T evaluated at the solution $\hat{\alpha}$. Hence, $\tilde{M}_T(\hat{\delta}) = o_p(1)$. Finally, the condition (iv) is easily proved by using Assumption (ii) and the identification condition $K_2 \geq G$. Therefore, we have $T^{1/2}(\hat{\alpha} - \alpha_0) = O_p(1)$ and

$$T^{1/2}(\hat{\alpha} - \alpha_0) = q Q_{zz}^{-1} H(\Pi_0)' M_T(\hat{\Delta}_0) + o_p(1),$$

which delivers the preliminary asymptotic representation[7] for the DSQR(θ, q):

$$T^{1/2}(\hat{\alpha} - \alpha_0) = Q_{zz}^{-1} H(\Pi_0)' \left\{ T^{-1/2} \sum_{t=1}^{T} q x_t \psi_\theta(v_t) \right.$$
$$\left. + (1-q) Q_0 T^{1/2}(\hat{\pi} - \pi_0) - Q_0 T^{1/2}(\hat{\Pi} - \Pi_0)\gamma_0 \right\} + o_p(1). \tag{A.9}$$

Substituting the asymptotic representations $T^{1/2}(\hat{\pi} - \pi_0) = Q_0^{-1} T^{-1/2} \sum_{t=1}^{T} x_t \psi_\theta(v_t) + o_p(1)$ and $T^{1/2}(\hat{\Pi}_j - \Pi_{0j}) = Q_j^{-1} T^{-1/2} \sum_{t=1}^{T} x_t \psi_\theta(V_{jt}) + o_p(1)$ into (A.9) shows that the first term involving q cancels out with the asymptotic representation of $T^{1/2}(\hat{\pi} - \pi_0)$ multiplied by $-qQ_0$. Simplifying gives

$$T^{1/2}(\hat{\alpha} - \alpha_0) = T^{-1/2} \sum_{t=1}^{T} Q_{zz}^{-1} H(\Pi_0)' x_t \psi_\theta(v_t)$$
$$- T^{-1/2} \sum_{i=1}^{G} \sum_{t=1}^{T} Q_{zz}^{-1} H(\Pi_0)' Q_0 Q_i^{-1} \gamma_{0i} x_t \psi_\theta(V_{it}) + o_p(1)$$
$$= D T^{-1/2} \sum_{t=1}^{T} Z_t + o_p(1) \tag{A.10}$$

where $Z_t = \psi_\theta(W_t) \otimes x_t$, $\psi_\theta(W_t) = (\psi_\theta(v_t), \psi_\theta(V_{1t}), \ldots, \psi_\theta(V_{Gt}))'$ and $D = Q_{zz}^{-1} H(\Pi_0)'[I, -Q_0 Q_1^{-1}\gamma_{01}, \ldots, -Q_0 Q_G^{-1}\gamma_{0G}]$. The expression in (A.10) is a scaled sample mean to which we can apply a CLT provided that $E(Z_t) = 0$, which is satisfied by Assumption 2(iv). Here, $E(Z_t) = 0$ is satisfied because the same quantile is used in both stages. Now, we examine what happens when LS estimation or a different quantile are used in the first stage. We first consider LS estimation. The corresponding asymptotic representations are $T^{1/2}(\hat{\pi}^{LS} - \pi_0) = Q^{-1} T^{-1/2} \sum_{t=1}^{T} x_t v_t + o_p(1)$ and $T^{1/2}(\hat{\Pi}^{LS} - \Pi_0)\gamma_0 = Q^{-1} T^{-1/2} \sum_{t=1}^{T} x_t V_t \gamma_0 + o_p(1)$, where $Q = E(x_t x_t')$. Substitution of these expressions into (A.19) gives $T^{1/2}(\hat{\alpha} - \alpha_0) = D^{LS} T^{-1/2} \sum_{t=1}^{T} Z_t^{LS} + o_p(1)$, where $D^{LS} = Q_{zz}^{-1} H(\Pi_0)'[qI, (1-q)Q_0 Q^{-1}, -Q_0 Q^{-1}]$ and $Z_t^{LS} = (x_t' \psi_\theta(v_t), x_t' v_t, x_t' V_t \gamma_0)'$. But, $E(Z_t^{LS}) \neq 0$ because $E\{\psi_\theta(v_t)|x_t\} = 0$ and $E\{v_t|x_t\} = 0$ do not generally hold simultaneously unless $\theta = 1/2$ and for symmetric distributions. Next, we investigate the consequence of using different quantiles (θ_1 for the first stage and θ_2 for the second stage). The asymptotic representations for the first-step estimators are given by $T^{1/2}(\hat{\pi}^Q - \pi_0) = Q_0^{-1} T^{-1/2} \sum_{t=1}^{T} x_t \psi_{\theta_1}(v_t) + o_p(1)$ and $T^{1/2}(\hat{\Pi}_j^Q - \Pi_{0j}) = Q_j^{-1} T^{-1/2} \sum_{t=1}^{T} x_t \psi_{\theta_1}(V_{jt}) + o_p(1)$. Plugging these representations into (A.9) results in $T^{1/2}(\hat{\alpha} - \alpha_0) = D^Q T^{-1/2} \sum_{t=1}^{T} Z_t^Q + o_p(1)$, where $D^Q = Q_{zz}^{-1} H(\Pi_0)'[qI, (1-q)I, -Q_0 Q_1^{-1}\gamma_{01}, \ldots, -Q_0 Q_G^{-1}\gamma_{0G}]$ and $Z_Q^t = (x_t' \psi_{\theta_2}(v_t), x_t' \psi_{\theta_1}(v_t), x_t' \psi_{\theta_1}(V_{1t}), \ldots, x_t' \psi_{\theta_1}(V_{Gt}))'$. Again, $E(Z_t^Q) \neq 0$ because $E\{\psi_{\theta_1}(v_t)|x_t\} = 0$ and $E\{\psi_{\theta_2}(v_t)|x_t\} = 0$ cannot hold unless $\theta_1 = \theta_2$. This issue of asymptotic bias caused by $E(Z_t) \neq 0$ is analysed in Kim and Muller (2000).

[7] An alternative approach would have been to work with an empirical process for the objective function itself: $\zeta_T(\Delta) = \sum_{t=1}^{T}[\rho_\theta(qv_t - T^{-1/2}x_t'\Delta) - \rho_\theta(qv_t)]$ and to use the method proposed in Pollard (1991).

B. APPENDIX. PROOFS

Proof of Proposition 1: Recalling Theorem 3.2 in Koenker and Bassett (1978), we have $\delta(\theta, \lambda y, X) = \lambda\delta(\theta, y, X)$ for any $\lambda > 0$ and $\delta(\theta, y + X\gamma, X) = \delta(\theta, y, X) + \gamma$, where γ is a parameter of appropriate dimension. The above invariance properties imply that $\delta(\theta, qy + (1 - q)XH(\hat{\Pi})\tilde{\alpha}, XH(\hat{\Pi})) = \delta(\theta, qy, XH(\hat{\Pi})) + (1 - q)\tilde{\alpha} = q\delta(\theta, y, XH(\hat{\Pi})) + (1 - q)\tilde{\alpha} = q\tilde{\alpha} + (1 - q)\tilde{\alpha} = \tilde{\alpha}$. This is the result (i). If $K_2 = G$ and $H(\hat{\Pi})$ is of full column rank, then $H(\hat{\Pi})$ is non-singular. Using $\delta(\theta, y, XA) = A^{-1}\delta(\theta, y, X)$ for any non-singular matrix A (Theorem 3.2 in Koenker and Bassett, 1978), we obtain $\tilde{\alpha} = \delta(\theta, y, XH(\hat{\Pi})) = H(\hat{\Pi})^{-1}\delta(\theta, y, X) = H(\hat{\Pi})^{-1}\hat{\pi}$. Next we can show that $\tilde{\alpha} = \hat{\alpha}$ because using (i) we have $\tilde{\alpha} = \delta(\theta, qy + (1 - q)XH(\hat{\Pi})\tilde{\alpha}, XH(\hat{\Pi})) = \delta(\theta, qy + (1 - q)X\hat{\pi}, XH(\hat{\Pi})) = \hat{\alpha}$. This shows (ii). $\square$

Proof of Proposition 2: Consider (A.10). Since Z_t is i.i.d. by Assumption 1, it is sufficient to show that $\text{var}(Z_t)$ is bounded to apply the Lindeberg–Levy's CLT. The moment condition on x_t in Assumption 2(i) is sufficient for this purpose because $\psi_\theta(\cdot)^2$ is bounded by 1. Noting that $\text{var}(Z_t) = \Omega$, we have $T^{-1/2}\sum_{t=1}^T Z_t \xrightarrow{d} N(0, \Omega)$, which proves the claim in the proposition. $\square$

Proof of Proposition 3: We first prove the claim $\hat{\Omega} \xrightarrow{p} \Omega$. Consider the $(1,1)$-submatrices of $\hat{\Omega}$ and Ω, which are given by $\hat{\Omega}_{11} = T^{-1}\sum_{t=1}^T \psi_\theta(\hat{v}_t)^2 x_t x_t'$ and $\Omega_{11} = E\{\psi_\theta(v_t)^2 x_t x_t'\}$. The consistency of $\hat{\Omega}_{11}$ is proved in two steps: (i) $|\Omega_{11T} - \Omega_{11}| = o_p(1)$ and (ii) $|\hat{\Omega}_{11} - \Omega_{11T}| = o_p(1)$ where $\Omega_{11T} = T^{-1}\sum_{t=1}^T \psi_\theta(v_t)^2 x_t x_t'$. The first step is obtained by applying the LLN for i.i.d. random variables under Assumptions 1 and 2(i). We now prove the second step. Consider the (i, j)-component of $|\hat{\Omega}_{11} - \Omega_{11T}|$ which is given by $|T^{-1}\sum_{t=1}^T\{\psi_\theta(\hat{v}_t) - \psi_\theta(v_t)\}\{\psi_\theta(\hat{v}_t) + \psi_\theta(v_t)\}x_{ti}x_{tj}| \le 2T^{-1}\sum_{t=1}^T |\psi_\theta(\hat{v}_t) - \psi_\theta(v_t)||x_{ti}||x_{tj}| \le 2T^{-1}\sum_{t=1}^T 1_{[|v_t|\le d_T]}|x_{ti}||x_{tj}|$, where $d_T = ||x_t|| \times ||\hat{\pi} - \pi_0||$. The first inequality comes from Minkowski's inequality and $|\psi_\theta(\cdot)| \le 1$ and the second inequality is obtained using $v_t - \hat{v}_t = x_t'(\hat{\pi} - \pi_0)$, $|x_t'(\hat{\pi} - \pi_0)| \le ||x_t|| \times ||\hat{\pi} - \pi_0||$ and the fact that $|1_{[x \le 0]} - 1_{[y \le 0]}| \le 1_{[|x| \le |x-y|]}$. Let $U_T = T^{-1}\sum_{t=1}^T 1_{[|v_t|\le d_T]}|x_{ti}||x_{tj}|$ and consider a set $A = \{U_T > \eta\}$ for $\eta > 0$. For any event B, we have $P(A) \le P(A \cap B) + P(B^c)$. We choose $B = \{||\hat{\pi} - \pi_0|| \le zT^{-d}\}$ where $z > 0$ and $0 < d < 1/2$. Then, we have $P(B^c) \to 0$ since $T^{1/2}(\hat{\pi} - \pi_0) = O_p(1)$. Now consider

$$P(A \cap B) \le (\eta T)^{-1}\sum_{t=1}^T E\left\{\int_{-||x_t||zT^{-d}}^{||x_t||zT^{-d}} f(\lambda|x_t)d\lambda|x_{ti}||x_{tj}|\right\}$$

(by the generalised Cebyshev inequality)

$$\le (\eta T)^{-1}\sum_{t=1}^T E\left\{\int_{-||x_t||zT^{-d}}^{||x_t||zT^{-d}} f_0 d\lambda|x_{ti}||x_{tj}|\right\}$$

(by Assumption 3(i))

$$= 2zf_0\eta^{-1}T^{-d}E\{||x_t|||x_{ti}||x_{tj}|\}.$$

The last expression converges to zero because $E\{||x_t|||x_{ti}||x_{tj}|\} < \infty$ by Assumption 2(i). Hence, we have proved that $U_T = T^{-1}\sum_{t=1}^T 1_{[|v_t|\le d_T]}|x_{ti}||x_{tj}| \xrightarrow{p} 0$ which in turn implies that $|\hat{\Omega}_{11} - \Omega_{11T}| = o_p(1)$. The second step is now proved. By combining (i) and (ii), we have $|\hat{\Omega}_{11} - \Omega_{11}| = o_p(1)$. The same argument can be applied to all other diagonal and off-diagonal terms of $\hat{\Omega}$ to show their consistency. Therefore, $|\hat{\Omega} - \Omega| = o_p(1)$.

Next, we turn to the claim $|\hat{D} - D| = o_p(1)$. We need to show the consistency of $\hat{Q}_0, \hat{Q}_1, \ldots, \hat{Q}_G, \hat{\Pi}$ and $\hat{\gamma}$. Since the results $\hat{\Pi} - \Pi_0 = o_p(1)$ and $\hat{\gamma} - \gamma_0 = o_p(1)$ are trivial, we first focus on $|\hat{Q}_0 - Q_0| = o_p(1)$. Let $Q_{0T} = (2c_{0T}T)^{-1}\sum_{t=1}^T 1_{[-c_{0T}\le v_t\le c_{0T}]}x_t x_t'$ and $\tilde{Q}_{0T} = (2c_{0T}T)^{-1}\sum_{t=1}^T 1_{[-\hat{c}_{0T}\le \hat{v}_t\le \hat{c}_{0T}]}x_t x_t'$. As before, the proof is carried out in three steps: (i) $|Q_{0T} - Q_0| = o_p(1)$, (ii) $|\tilde{Q}_{0T} - Q_{0T}| = o_p(1)$ and (iii) $|\tilde{Q}_{0T} - \hat{Q}_{0T}| = o_p(1)$.

We start with (i). The Mean value theorem implies that $E(Q_{0T}) = E\{T^{-1}\sum_{t=1}^{T} f(\xi_T|x_t)x_t x_t'\}$, where $-c_{0T} \leq \xi_T \leq c_{0T}$. Noting that $\xi_T = o_p(1)$, $|E(Q_{0T}) - Q_0| = o(1)$ by Minkowski inequality and Assumptions 2(i) and 2(ii). Using an LLN for double arrays, we have $|Q_{0T} - E(Q_{0T})| = o_p(1)$. Therefore, the first step is proven.

Now we turn to (ii). Using the fact that $|1_{[x \leq 0]} - 1_{[y \leq 0]}| \leq 1_{[|x| \leq |x-y|]}$, the (i, j)th element of $|\tilde{Q}_{0T} - Q_{0T}|$ given by $|(2c_{0T}T)^{-1}\sum_{t=1}^{T}(1_{[-\hat{c}_{0T} \leq \hat{v}_t \leq \hat{c}_{0T}]} - 1_{[-c_{0T} \leq v_t \leq c_{0T}]})x_t x_t'|$ is bounded by $(2c_{0T}T)^{-1}\sum_{t=1}^{T}$ $(1_{[|v_t+c_{0T}| \leq d_T]} + 1_{[|v_t-c_{0T}| \leq d_T]})|x_{ti}||x_{tj}| = U_{1T} + U_{2T}$, where $d_T = ||x_t|| \times ||\hat{\pi} - \pi_0|| + |\hat{c}_{0T} - c_{0T}|$, $U_{1T} = (2c_{0T}T)^{-1}\sum_{t=1}^{T} 1_{[|v_t+c_{0T}| \leq d_T]}|x_{ti}||x_{tj}|$ and $U_{2T} = (2c_{0T}T)^{-1}\sum_{t=1}^{T} 1_{[|v_t-c_{0T}| \leq d_T]}|x_{ti}||x_{tj}|$. By using the same argument used to show $U_T \to 0$ in the proof of $|\hat{\Omega}_{11} - \Omega_{11T}| = o_p(1)$, one can show $U_{1T} = o_p(1)$ and $U_{2T} = o_p(1)$, which implies $|\tilde{Q}_0 - Q_0| = o_p(1)$. The second step is proven.

To show step (iii), we note that $\hat{Q}_{0T} - \tilde{Q}_{0T} = (c_{0T}/\hat{c}_{0T} - 1)\tilde{Q}_{0T}$. Since $\tilde{Q}_{0T} = O_p(1)$ and $(c_{0T}/\hat{c}_{0T} - 1) = o_p(1)$ by Assumption 3(ii), the last step is proved. Therefore, we have: $|\hat{Q}_0 - Q_0| = o_p(1)$. The same argument can be used to show $|\hat{Q}_j - Q_j| = o_p(1)$ for $j = 1, \ldots, G$. Therefore, we have $\hat{D} \xrightarrow{p} D$.   □

C. APPENDIX. SIMULATION DESIGN

The structural system is given by $B[\begin{smallmatrix} y_t' \\ Y_t' \end{smallmatrix}] + \Gamma x_t' = U_t'$, where $[\begin{smallmatrix} y_t' \\ Y_t' \end{smallmatrix}]$ is a 2×1 vector of endogenous variables, x_t' is a 4×1 vector of exogenous variables with the first element set to one, U_t' is a 2×1 vector of errors,

$$B = \begin{bmatrix} 1 & -0.5 \\ -0.7 & 1 \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & -0.5 \end{bmatrix}.$$

We are interested in the first equation of the system and the system is over-identified by the zero restrictions $\Gamma_{13} = \Gamma_{14} = \Gamma_{22} = 0$. Here, the parameters in (1) are $\gamma_0 = 0.5$ and $\beta_0' = (1, 0.2)$, X_1 is the first two columns in X and u is the first column in U. The above structural equation can be written as $[yY]B' = -X\Gamma' + U$, which gives the following reduced form equations $[yY] = X[\pi_0\Pi_0] + [vV]$, where $[\pi_0\Pi_0] = -\Gamma'(B')^{-1}$ and $[vV] = U(B')^{-1}$. We obtain $\pi_0' = (2.3, 0.3, 0.3, -0.15)$ and $\Pi_0' = (2.6, 0.2, 0.6, -0.3)$.

The errors $[vV]$ in the reduced form equations are generated so that Assumption 2(iv) is satisfied: $v = v^e - F^{v^e-1}(\theta)$ and $V = V^e - F^{V^e-1}(\theta)$, where v^e and V^e are generated for the different simulation sets by using the three distributions N(0, 1), $t(3)$ and LN(0,1) with correlation -0.1, and $F^{v^e-1}(\theta)$ and $F^{V^e-1}(\theta)$ are the inverse cumulative functions of v^e and V^e evaluated at θ. The second to fourth columns in X are generated using the normal distribution with zero means and covariances, and unit variances. Finally, we generate the endogenous variables $[yY]$ using the reduced-form equations.

AUTHOR QUERIES

Journal: ECTJ
Paper: ectj˙001

Dear Author

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Query Reference	Query	Remarks
Q1	Author: Please check, Table 2(a) and 2(b) have been changed to Table 2 and 3.	
Q2	Author: If possible please provide more information of Reference Cowell and Flachaire, 2000.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	⋯ under matter to remain	Stet
Insert in text the matter indicated in the margin	⅄	New matter followed by ⅄
Delete	⊢⊣ through matter to be deleted	ℐ
Delete and close up	⊕ through matter to be deleted	ℐ
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢⊣ through word	New letter or new word
Change to italics	— under matter to be changed	⊔⊔⌡
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	═ under matter to be changed	═
Change to bold type	∿ under matter to be changed	∿
Change to bold italic	≈ under matter to be changed	⊔⊔⌡
Change to lower case	Encircle matter to be changed	≢
Change italic to upright type	(As above)	⊔⌡
Insert 'superior' character	/ through character or ⅄ where required	⅄ under character e.g. ⅄
Insert 'inferior' character	(As above)	⅄ over character e.g. ⅄₂
Insert full stop	(As above)	⊙
Insert comma	(As above)	﹐
Insert single quotation marks	(As above)	⅄ and/or ⅄
Insert double quotation marks	(As above)	⅄ and/or ⅄
Insert hyphen	(As above)	⊞
Start new paragraph	⌐	⌐
No new paragraph	↰	↰
Transpose	⊔⌐	⊔⌐
Close up	linking ⊃ letters	⊃
Insert space between letters	⅄ between letters affected	#
Insert space between words	⅄ between words affected	#
Reduce space between letters	↑ between letters affected	↑
Reduce space between words	↑ between words affected	↑