UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/14

TWO STAGE QUANTILE REGRESSION WHEN THE FIRST STAGE IS BASED ON QUANTILE REGRESSION

by Tae-Hwan Kim and Christophe Muller

August 2003 DP 03/14

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 03/14

TWO-STAGE QUANTILE REGRESSION WHEN THE FIRST STAGE IS BASED ON QUANTILE REGRESSION

by Tae-Hwan Kim and Christophe Muller

Tae-Hwan Kim is Lecturer and Christophe Muller is CREDIT Research Fellow, both in the School of Economics, University of Nottingham

August 2003

Two-stage quantile regression

when the first stage is based on quantile regression

TAE-HWAN KIM, CHRISTOPHE MULLER

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

Web address: http://www.nottingham.ac.uk/economics/

E-mail: Tae-Hwan.Kim@nottingham.ac.uk
E-mail: Christophe.Muller@nottingham.ac.uk

First submitted March 2001.

May 2003

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

Applied researchers increasingly rely on quantile regressions for their model estimations. This method is often used for wage and living standard analyses[1], but also for studies of firm data[2] and financial data[3], as well as in non-economic domains[4]. 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 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, socio-economic 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) deal 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 Amemiya and Powell works by using quantile regressions and random exogenous variables[5]. We use the same quantile estimation in the two steps and the resulting estimator of the structural parameter is termed "Double-Stage Quantile Regression estimator (DSQR)."

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) study least absolute error difference estimators (LAED) for estimating a single equation from a simultaneous equation model[6]. Abadie, Angrist and Imbens (2002) design a quantile

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

[2] Mata and Machado (1996), Machado and Mata (2000).

[3] Engle and Manganelli (1999).

[4] Lipsitz et al. (1997).

[5] Koenker and Bassett (1978, 82), Bassett and Koenker (1978, 86), Powell (1983).

[6] 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.

treatment effects estimator (QTE), which is the solution to a convex programming problem with first-step nonparametric estimation of a nuisance function[7]. MaCurdy and Timmins (2000) propose an estimator for ARMA models adapted to the quantile regression framework. [8]

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[9] 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 shall focus on the case when the first stage is composed of quantile regressions where the quantile is the same as for the second stage.

In this paper, we study the asymptotic properties of the DSQR and we provide invariance results related to the reformulation of the dependent variable. We also propose a consistent estimator of the asymptotic variance-covariance matrix of the DSQR. Finally, we investigate the finite sample properties of this estimator. We emphasize that we study all these properties in the case of random regressors, in contrast with Amemiya (1982) and Powell (1983) in which all exogenous variables are assumed fixed.

What are the asymptotic, finite sample, invariance and robustness properties of the DSQR? How can we estimate its precision? The aim of this paper is to study these questions. 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 present simulation results in Section 4. Finally, Section 5 concludes. The proofs are presented in Appendix 2.

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

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

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

[9] Arias, Hallock and Sosa-Escudero (2001), Garcia, Hernandez and Lopez (2001).

$K_2(\equiv K - K_1)$ exogenous variables absent from (1). In this situation the endogeneity of Y in (1) may cause that $Q_\theta(u|Y) \neq Q_\theta(u)$, 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. We use this non-equality as a formal 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 $X \equiv [X_1, X_2]$ 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 $\pi_0 \equiv \left[\Pi_0, \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}\right] \alpha_0 \equiv H(\Pi_0)\alpha_0$ and $v \equiv u + V\gamma_0$.

Define the "check function" $\rho_\theta : R \to R^+$ for given $\theta \in (0,1)$ as

$$\rho_\theta(z) \equiv z\psi_\theta(z),$$

where $\psi_\theta(z) \equiv \theta - 1_{[z \leq 0]}$ 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:

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

where y_t and x_t' are the t^{th} elements in y and X respectively, q is a non-zero constant chosen in advance by the researcher, and where $\hat{\pi}$ and $\hat{\Pi}_j$ 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 is a T-vector a parameters, Π_j and Y_j are respectively the j^{th} columns of the $K \times G$ matrix of parameters Π and 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 $\beta(\theta, y, X)$ denotes the quantile regression estimator associated with quantile θ, dependent variable y and matrix of right-hand-side variables X. Moreover, let $\tilde{\alpha} \equiv \beta(\theta, y, XH(\hat{\Pi}))$. Then, we have for $q > 0$:*

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

(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 shall 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. Since the choice of q generally intervenes in the estimation, optimal values of q can be searched to improve the estimation. However, we shall show in Section 3 that the asymptotic distribution of the DSQR(θ, q) does not depend on q. Moreover, (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 here is analogous to that used for Indirect Least-Squares estimator (ILS). Indeed, (ii) shows that estimates of the coefficient in regressions of reduced-form equations can be used to calculate structural coefficients, but only if there is one-to-one correspondance between the structural parameters and the reduced form. Proposition 1 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 the precisely same way in which the ILS is obtained.

In the next section we discuss the asymptotic representation of the DSQR(θ, q). The following conditions will be useful to obtain this asymptotic representation and the asymptotic normality of the DSQR(θ, q).

Assumption 1 *The sequence $\{(u_t, V_t, x_t)\}$ 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) $E(\|x_t\|^4) < \infty$, where $\|x_t\| = (x_t' x_t)^{1/2}$.
(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, $Q_0 \equiv E\{f(0|x_t)x_t x_t'\}$ and $Q_j \equiv E\{g_j(0|x_t)x_t x_t'\}$ are positive definite.*
(iv) *$E\{\psi_\theta(v_t) \mid x_t\} = 0$ and $E\{\psi_\theta(V_{jt}) \mid x_t\} = 0$ $(j = 1, \ldots, G)$.*

Assumption 2(i) is needed to generalise the stochastic equicontinuity result in Powell (1983) and to prove the consistency of our covariance matrix estimator. Assumption 2(iv) 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. When there is an intercept term in the model, Assumption 2(iv) is an identification condition on the coefficient of the intercept.

Assumption 2 (iv) normalises the intercept on the θ^{th} 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. The problem is that for $q \neq 1$ distinct methods for first and second stages generally imply incompatible restrictions on error terms, except in restrictive cases. This is the case when different quantiles are used for the first and second stages since it would entail that two different conditional quantiles of v_t are null. Then, this contradiction is resolved by the occurrence of a bias in the intercept coefficient.

We also note that the normalisation in Assumption 2 (iv) systematically simplifies the asymptotic representation by enabling us to substitute $f_{v|X}(F_{v|X}^{-1}(\theta))$ with $f_{v|X}(0)$. This is an additional reason to favour this restriction, which facilitates the estimation of the accuracy of the estimators since we need to estimate nonparametrically $f_{v|X}^{-1}$ at 0 instead of $f_{v|X}$ and $F_{v|X}^{-1}$ together. 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) estimator is based on its asymptotic representation derived in Appendix A. It is easy to see from this representation that the DSQR(θ, q) estimator is asymptotically robust because its influence function is a linear combination of bounded functions. However, the robustness would be lost if non-robust first-stage estimators were used instead of quantile regressions. Therefore, one major advantage of the DSQR(θ, q) is that it preserves the robustness of quantile regressions.

Proposition 2 *Under Assumptions 1 and 2,* $T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, D\Omega D')$, *where* $D = Q_{zz}^{-1} H(\Pi_0)'[I, -Q_0 Q_1^{-1}\gamma_{01}, ..., -Q_0 Q_G^{-1}\gamma_{0G}]$, $Q_{zz} = H(\Pi_0)' Q_0 H(\Pi_0)$ *and* $\Omega = E(\Sigma \otimes x_t x_t')$, *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 representation. However, $\hat{\pi}$ can still be used for a consistent estimator of the asymptotic covariance matrix. We derive estimators of D and Ω by using the plug-in principle:

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

where $\hat{Q}_{zz} = H(\hat{\Pi})'\hat{Q}_0 H(\hat{\Pi})$, $\hat{Q}_0 = (2\hat{c}_{0T}T)^{-1}\sum_{t=1}^{T} 1_{[-\hat{c}_{0T}\leq\hat{v}_t\leq\hat{c}_{0T}]}x_t x_t'$,

$\hat{Q}_j = (2\hat{c}_{jT}T)^{-1}\sum_{t=1}^{T} 1_{[-\hat{c}_{jT}\leq\hat{V}_{jt}\leq\hat{c}_{jT}]}x_t x_t'$, $\hat{v}_t = y_t - x_t'\hat{\pi}$, $\hat{V}_{jt} = Y_{jt} - x_t'\hat{\Pi}$ and $\hat{c}_{jT}$ $(j = 1, 2, ..., G)$ are data-dependent bandwidth;

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

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, the covariance matrix estimator is $\hat{D}\hat{\Omega}\hat{D}'$. For the consistency of the covariance matrix estimator we need the following additional assumptions.

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 $j = 0, 1, ..., G$ such that we have $\hat{c}_{jT}/c_{jT} \xrightarrow{p} 1$, $c_{jT} = o_p(1)$ and $c_{jT}^{-1} = o_p(T^{1/2})$.*

Proposition 3 *Under Assumptions 1,2 and 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, including the case of DSLAD and even with non-random regressors. Note that directly substituting consistent estimators in the formula of the covariance matrix in Proposition 2 generally produces a less efficient estimator of the variance-covariance matrix when the following simplifications on diagonal terms are ignored: $E\left[\{\psi_\theta(W_{it})\}^2 x_t x_t'\right] = \theta(1-\theta)E\left[x_t x_t'\right]$. Alternatively, one can omit the first-stage estimation with $\hat{\pi}$ by using $\tilde{v}_t = \hat{u}_t + \hat{V}_t\hat{\gamma}$ instead of $\hat{v}_t$. 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)[10]. Here, parameter q affects the empirical distribution of the estimator, in contrast with asymptotic results. We are interested in comparing the small sample behaviour of the quantile estimator for the structural parameters (γ_0, β_0) in the two cases: (1) when the endogeneity problem is ignored and (2) when the problem is corrected by using our procedure. We also want to examine the robustness of the DSQR(θ, q) in small samples.

[10] See also Ribeiro (1998) for small sample simulations of LAD estimators with a first step of LS or LAD estimators.

The data generating process used in the simulations is described in Appendix C. The first (structural) equation has two endogenous variables and two exogenous variables including a constant, while the total number of exogenous variables in the system is 4. The first equation is over-identified by construction and corresponds to parameter values: $\gamma_0 = 0.5$ 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 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) based on the 1,000 replications.

The performance of the one-stage quantile regression estimator for the different distributions is displayed in Tables 1-3. This estimator exhibits a systematic bias in finite samples, which does not disappear as the sample size increases. The results vary little when increasing the number of observations from 50 to 300.

The results for the DSQR(θ, q), denoted $(\hat{\gamma}\ , \hat{\beta})$, are provided in Tables 4-6. We select three values (0.1, 0.5 and 1) for q. As shown in Proposition 1, when the system is exactly identified, this dependence on q disappears and we confirm it by 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. These features for the effect of q are also obtained with $t(3)$ and LN$(0, 1)$.

We now discuss the results of Tables 4(a)-4(c) with normal errors. 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. The results with Student-t error terms, which are partly displayed in Table 5 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, partly shown in Table 6, differ in that both estimators are severely biased for large quantiles (for $\theta = 0.95$ and LN(0,1)). 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) is the best for small quantiles (eg, $\theta = 0.05$), 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, the mode of the distribution can be far from zero. Then $g_j(0|x_t)$ may be smaller than for a symmetric distribution. This inflates the roles of some terms in the diagonal terms of $D\Omega D'$, in particular the ones including $[E\{g_j(0|x_t)x_t x_t'\}]^{-2}$. In particular, off diagonal terms of Ω involve a factor θ^2 and contribute to the diagonal terms of $D\Omega D'$. Then, these terms may have larger absolute magnitudes for θ close to 1, and smaller for θ close to 0. This may cause the large bias for $\theta = 0.95$ and LN(0,1). On the whole, the DSQR(θ, q) well tackles the extreme values that occur more frequently with Student-t and Lognormal distributions.

For all types of error terms, the one-stage estimator is severely biased with small samples. In contrast, the DSQR(θ, q) has good finite sample properties, although a too small sample size or extreme quantiles ($\theta = 0.05, 0.95$) may degrade its performance. We have calculated some robust measures (sample medians and sample interquartile ranges) 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 Tables 4. The dispersion of the sampling distribution of the deviations is slightly smaller when robust measures are used.

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. Following Cowell and Flachaire (2002), we generate one outlier in each replication by randomly selecting one observation and multiplying it by 15. The results are reported in Table 7 for the 2SLS, which is invariant to the value of q, and for the DSQR$(\theta, 1)$. We report only for the normal error case and $q = 1$ because the 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 the robust measures are used (medians and interquartiles indicators).

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 based 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 ancillary predictive quantile regressions.

ACKNOWLEDGEMENTS

First submitted in December 2001. We are grateful to R. Davidson, W. Newey, J. Powell, H. White and O. Linton for useful discussions about a preliminary working paper and to participants in presentations in Nottingham University, CREST-INSEE in Paris, London School of Economics, University of California at San Diego, the 1999 Econometric Study Group Meeting in Bristol, the 2000 Royal Statistical Society Conference in Reading, and two referees and an editor for their comment. Remaining errors are ours. We acknowledge a grant (no. 30,649) from the British Academy.

References

[1] 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*, 91-118.

[2] Amemiya, T. (1982). Two stage least absolute deviation estimators. *Econometrica* 50, 689-711.

[3] Andrews, D.W.K. (1990). Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity and nonparametric kernel estimation. Cowles Foundation for Research in Economics, Yale University, July.

[4] 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.

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

[6] Chen, L.-A. and S. Portnoy (1996). Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models. *Commun. Statist.-Theory Meth.* 25, 1005-32.

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

[8] Engle, R.F. and S. Manganelli (1999). CAViaR: Conditional autoregressive Value at Risk by regression quantiles. mimeo University of California-San Diego.

[9] 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.

[10] Jureckova, J. (1977). Asymptotic relations of M-Estimates and R-Estimates in linear regression models. *The Annals of Statistics* 5, 464-72.

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

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

[13] Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33-50.

[14] Lipsitz, S.R., G.M. Fitzmaurice, G. Molenberghs and L.P. Zhao (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immonodeficiency virus. *Applied Statistics* 46, 463-76.

[15] Machado, A.F. and J. Mata (2000). Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* 15, 253-74.

[16] Machado, A.F. and J. Mata (2001). Earnings functions in Portugal 1982-1994: Evidence from quantile regressions. *Empirical Economics* 26, 115-34.

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

[18] Mata, J. and J.A.F. Machado (1996). Firm start-up size: A conditional quantile approach. *European Economic Review* 40, 1305-23.

[19] Newey, W.K. (1985). Semiparametric estimation of limited dependent variables models with endogenous explanatory variables. *Annales de l'INSEE*, n. 59/60, 219-35.

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

[21] Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303-25.

[22] Powell, J. (1985). Censored regression quantiles. *Journal of Econometrics* 32, 143-55.

[23] Ribeiro, E. (1998). Small sample evidence of quantile regression for structural models: estimation and testing. *Revista de Econometria* 18.

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

A. APPENDIX. ASYMPTOTIC REPRESENTATION

We define an empirical process $M_T(\cdot)$ by

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

where Δ is a $K\times 1$ vector, $w_t = (v_t, x_t')'$ and $m(w_t,\Delta) = x_t\psi_\theta(qv_t - T^{-1/2}x_t'\Delta)$. Since the funtion ψ_θ is of bounded variations, the i.i.d. condition in Assumption 1 and the moment condition on x_t in Assumption 2(i) are sufficient to apply Theorems I.1 and I.3 in Andrews (1990), which leads to:

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

This result is a generalisation of the lemma in Powell (1983) when x_t is random.

We outline the proof below. Define $V_T(\Delta) \equiv T^{-1/2}\sum_{t=1}^{T}[m(w_t,\Delta) - E\{m(w_t,\ \Delta)\}] = M_T(\Delta) - E(M_T(\Delta))$. Then, we have

$$\sup_{||\Delta_1-\Delta_2||\leq L} ||V_T(\Delta_1) - V_T(\Delta_2)|| = o_p(1). \tag{6}$$

The claim in 6 is a direct application of Theorem II.1 in Andrews (1990), but in order to apply the theorem the following two conditions must be verified:
(i) $m(\cdot,\Delta)$ satisfies Pollard's entropy condition with some envelop $\bar{M}(w_t)$.
(ii) For some $\delta > 2$, $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[\{\bar{M}(w_t)\}^{\delta}\right] < \infty$.

Let $f_1(w,\Delta) = x$ and $f_2(w,\Delta) = \psi_\theta(qv - T^{-1/2}x'\Delta)$ so that $m(\cdot,\Delta) = f_1(\cdot,\Delta)f_2(\cdot,\Delta)$. Since $f_1(\cdot,\Delta)$ is a Type I function and $f_2(\cdot,\Delta)$ is a Type II function with Lipschitz coefficient $B(w) = ||x||$ (see Andrews, 1990, for definitions). $m(\cdot,\Delta)$ satisfies Pollard's entropy condition with envelope $||x||$ because it is a product of Type I & II functions (See Theorem II.3 in Andrews, 1990). Hence, (i) is verified. The second condition is now given by $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[\{\bar{M}(w_t)\}^{\delta}\right] = E\left\{||x_t||^{\delta}\right\}$ which is bounded by Assumption 2(i).

The result in (6) can be written by setting $\Delta_1 = \Delta$ and $\Delta_2 = 0$ as

$$\sup_{||\Delta||\leq L} ||M_T(\Delta) - M_T(0) - \{E(M_T(\Delta)) - E(M_T(0))\}|| = o_p(1). \tag{7}$$

Now we want to show that $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$ as follows. Noting that $E(M_T(\Delta)) = E\left\{T^{-1/2}\sum_{t=1}^{T}\left[x_t\theta - x_t\int_{-\infty}^{q^{-1}x_t'T^{-1/2}\Delta} f(v|x_t)dv\right]\right\}$, we have $E(M_T(\Delta)) - E(M_T(0))$

$$\begin{aligned} &= -E\left\{T^{-1/2}\sum_{t=1}^{T}\left[x_t\int_{0}^{q^{-1}x_t'T^{-1/2}\Delta} f(v|x_t)dv\right]\right\} \\ &= -E\left\{q^{-1}T^{-1}\sum_{t=1}^{T} x_t x_t'\Delta\frac{F(q^{-1}x_t'T^{-1/2}\Delta|x_t) - F(0|x_t)}{q^{-1}x_t'T^{-1/2}\Delta}\right\}\end{aligned}$$

where $F(\cdot|x_t)$ is the conditional cdf. Let $G(\lambda) = q^{-1}T^{-1}\sum_{t=1}^{T}F(\lambda|x_t)x_tx_t'\Delta$. Then, by the Mean Value Theorem and the continuity in Assumption 2, there exist ξ_T between 0 and $q^{-1}x_t'T^{-1/2}\Delta$ such that $E(M_T(\Delta)) - E(M_T(0)) = -E\left\{G'(\xi_T)\right\} = -q^{-1}E\left\{T^{-1}\sum_{t=1}^{T}f(\xi_T|x_t)x_tx_t'\right\}\Delta \to -q^{-1}Q_0\Delta$,where $Q_0 = E\left\{f(0|x_t)x_tx_t'\right\}$. We now show the latter convergence result.

Let $Q_T \equiv T^{-1}\sum\limits_{t=1}^{T} f(\xi_T|x_t)x_tx_t'$ and consider the $(i,j)^{th}$ element of $|E(Q_T) - Q_0|$, which is given by $\left|T^{-1}\sum\limits_{t=1}^{T} E\left(\left\{f(\xi_T|x_t) - f(0|x_t)\right\}x_{ti}x_{tj}'\right)\right| \le$

$T^{-1}\sum\limits_{t=1}^{T} E\left(\left|f(\xi_T|x_t) - f(0|x_t)\right| \, |x_{ti}| \, |x_{tj}|\right)$ (by Minkowski's inequality and Jensen's inequality).

$\le T^{-1}\sum\limits_{t=1}^{T} E\left(L_0\,|\xi_T| \, |x_{ti}| \, |x_{tj}|\right)$ (for some constant L_0 by Assumption 2(ii))

$\le L_0\,E(\xi_T^2)^{1/2}T^{-1}\sum\limits_{t=1}^{T} E\left(x_{ti}^2\,x_{tj}^2\right)^{1/2}$. Note that $0 \le E(\xi_T^2) \le \frac{1}{q^2}\frac{1}{T}E(x_t'\Delta)^2 \longrightarrow 0$, which implies $E(\xi_T^2) \to 0$. Since $\frac{1}{T}\sum\limits_{t=1}^{T} E\left(x_{ti}^2\,x_{tj}^2\right)^{1/2} = O(1)$, we have $|E(Q_T) - Q_0| \xrightarrow{p} 0$, i.e. $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$.

Hence, we now can replace $E(M_T(\Delta)) - E(M_T(0))$ in (7) with its limit $-q^{-1}Q_0\Delta$ to obtain

$$\sup_{||\Delta||\le L} ||M_T(\Delta) - M_T(0) + q^{-1}Q_0\Delta|| = o_p(1).$$

Combining the result and the fact that $T^{1/2}(\hat{\pi} - \pi_0) = O_p(1)$ and $T^{1/2}(\hat{\Pi} - \Pi_0) = O_p(1)$, we obtain the asymptotic representation for the DSQR(θ, q):

$$\begin{aligned} T^{1/2}(\hat{\alpha} - \alpha_0) &= Q_{zz}^{-1}H(\Pi_0)'\{T^{-1/2}\sum_{t=1}^{T} qx_t\psi_\theta(v_t) \qquad (8)\\ &\quad +(1-q)Q_0T^{1/2}(\hat{\pi} - \pi_0) - Q_0T^{1/2}(\hat{\Pi} - \Pi_0)\gamma_0\} + o_p(1)\end{aligned}$$

where $Q_{zz} = H(\Pi_0)'Q_0 H(\Pi_0)$. 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 (8) shows that the first term involving q cancels out with the asymptotic representation of $T^{1/2}(\hat{\pi} - \pi_0)$ multiplied by $-qQ_0$. Collecting terms and simplifying them gives $T^{1/2}(\hat{\alpha} - \alpha_0)$

$$
\begin{aligned}
&= T^{-1/2}\sum_{t=1}^{T} Q_{zz}^{-1}H(\Pi_0)'x_t\psi_\theta(v_t) + o_p(1) \\
&\quad -T^{-1/2}\sum_{t=1}^{T} Q_{zz}^{-1}H(\Pi_0)'Q_0Q_1^{-1}\gamma_{01}x_t\psi_\theta(V_{1t}) + o_p(1) \\
&\quad \cdots \\
&\quad -T^{-1/2}\sum_{t=1}^{T} Q_{zz}^{-1}H(\Pi_0)'Q_0Q_G^{-1}\gamma_{0G}x_t\psi_\theta(V_{Gt}) + o_p(1) \\
&= DT^{-1/2}\sum_{t=1}^{T} Z_t + o_p(1) \qquad (9)
\end{aligned}
$$

where $Z_t = \psi_\theta(W_t)'\otimes x_t$ and $D = Q_{zz}^{-1}H(\Pi_0)'[I, -Q_0Q_1^{-1}\gamma_{01}, ..., -Q_0Q_G^{-1}\gamma_{0G}]$. The expression in (9) 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). In particular, $E(Z_t) = 0$ is met since the same quantile is used in both stages. Now, we can examine what happens when LS estimation or different quantiles are used in the first stage. We consider LS estimation first. 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 (8) gives $T^{1/2}(\hat{\alpha} - \alpha_0) = D^{LS}T^{-1/2}\sum_{t=1}^{T} Z_t^{LS} + o_p(1)$, where $D^{LS} = Q_{zz}^{-1}H(\Pi_0)'[qI, (1-q)Q_0Q^{-1}, -Q_0Q^{-1}]$ and $Z_t^{LS} = (x_t'\psi_\theta(v_t), x_t'v_t,\ x_t'V_t\gamma_0)'$. It is clear from this expression that $E(Z_t^{LS}) \neq 0$ 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 (8) 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 $D^Q = Q_{zz}^{-1}H(\Pi_0)'[qI, (1-q)I, -Q_0Q_1^{-1}\gamma_{01}, ..., -Q_0Q_G^{-1}\gamma_{0G}]$ and $Z_t^Q = (x_t'\psi_{\theta_2}(v_t), x_t'\psi_{\theta_1}(v_t), x_t'\psi_{\theta_1}(V_{1t}), ...,\ x_t'\psi_{\theta_1}(V_{Gt}))'$. Again, we see that $E(Z_t^Q) \neq 0$ 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 depth in Kim and Muller (2000). They show that the bias can be isolated in the intercept coefficient, provide an explicit expression of the bias and derive the limiting distribution of the slope coefficients.

B. APPENDIX. PROOFS

Proof of Proposition 1: Recalling Theorem 3.2 in Koenker and Bassett (1978), we have $\beta(\theta, \lambda y, X) = \lambda\beta(\theta, y, X)$ for any $\lambda > 0$ and $\beta(\theta, y + X\gamma, X) = \beta(\theta, y, X) + \gamma$, where γ is a parameter of appropriate dimension. The above invariance properties imply that $\beta(\theta, qy+(1-q)XH(\hat{\Pi})\tilde{\alpha}, XH(\hat{\Pi}))$ $= \beta(\theta, qy, X\ H(\hat{\Pi}))+(1-q)\tilde{\alpha} = q\beta(\theta, y, XH(\hat{\Pi}))+(1-q)\tilde{\alpha} = q\tilde{\alpha}+(1-q)\tilde{\alpha} = \tilde{\alpha}$. This is result (i).

If $K_2 = G$ and $H(\hat{\Pi})$ is of full column rank, then $H(\hat{\Pi})$ is non singular. Using $\beta(\theta, y, XA) = A^{-1}\beta(\theta, y, X)$ for any non singular matrix A (Theorem 3.2 in Koenker and Bassett, 1978), we obtain $\tilde{\alpha} = \beta(\theta, y, XH(\hat{\Pi})) = H(\hat{\Pi})^{-1}\beta(\theta, y, X) = H(\hat{\Pi})^{-1}\hat{\pi}$. Next we can show that $\tilde{\alpha} = \hat{\alpha}$ because using (i) we have $\tilde{\alpha} = \beta(\theta, qy + (1-q)XH(\hat{\Pi})\tilde{\alpha}, XH(\hat{\Pi})) = \beta(\theta, qy + (1-q)X\hat{\pi}, XH(\hat{\Pi})) = \hat{\alpha}$. This shows (ii). *QED.*

Proof of Proposition 2: Consider (9). Since Z_t is i.i.d. by Assumption 1, it is sufficient to show that $var(Z_t)$ 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 $var(Z_t) = \Omega$, we have $T^{-1/2}\sum_{t=1}^{T} Z_t \xrightarrow{d} N(0, \Omega)$, which proves the claim in the proposition. *QED.*

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) $|\Omega_{11T} - Q_{11}| = o_p(1)$ 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 a straightforward application of 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

$$
\begin{aligned}
& |T^{-1}\sum_{t=1}^{T}\{\psi_\theta(\hat{v}_t) - \psi_\theta(v_t)\}\{\psi_\theta(\hat{v}_t) + \psi_\theta(v_t)\}x_{ti}x_{tj}| \\
\leq \quad & 2T^{-1}\sum_{t=1}^{T}|\psi_\theta(\hat{v}_t) - \psi_\theta(v_t)||x_{ti}||x_{tj}| \\
\leq \quad & 2T^{-1}\sum_{t=1}^{T} 1_{[|v_t| \leq d_T]}|x_{ti}||x_{tj}|
\end{aligned}
$$

where $d_T = ||x_t|| \times ||\hat{\pi} - \pi_0||$. The first inequality is due to Minkowski's inequality and $|\psi_\theta(\cdot)| \leq 1$ and the second inequality is obtained using $v_t - \hat{v}_t = x_t'(\hat{\pi} - \pi_0)$, $|x_t'(\hat{\pi} - \pi_0)| \leq ||x_t|| \times ||\hat{\pi} - \pi_0||$ and the fact that $|1_{[x\leq 0]} - 1_{[y \leq 0]}| \leq 1_{[|x| \leq |x-y|]}$. Let $U_T = T^{-1}\sum_{t=1}^{T} 1_{[|v_t| \leq d_T]}|x_{ti}||x_{tj}|$ and consider a set $A = \{U_T > \eta\}$ for $\eta > 0$. For any event B, we have $P(A) \leq P(A \cap B) + P(B^c)$. We choose $B = \{||\hat{\pi} - \pi_0|| \leq zT^{-d}\}$ 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

$$\begin{aligned} P(A \cap B) &\leq (\eta T)^{-1} \sum_{t=1}^{T} E\left\{ \int_{-||x_t||zT^{-d}}^{||x_t||zT^{-d}} f(\lambda | x_t) d\lambda |x_{ti}||x_{tj}| \right\} \\ &\qquad \text{(by the generalised Cebyshev inequality)} \\ &\leq (\eta T)^{-1} \sum_{t=1}^{T} E\left\{ \int_{-||x_t||zT^{-d}}^{||x_t||zT^{-d}} f_0 d\lambda |x_{ti}||x_{tj}| \right\} \\ &\qquad \text{(by Assumption 3(i))} \\ &= 2zf_0\eta^{-1}T^{-d}E\left\{||x_t|||x_{ti}||x_{tj}|\right\} \end{aligned}$$

The last expression converges to zero because $E\{||x_t|||x_{ti}||x_{tj}|\} < \infty$ by Assumption 2(i). Hence, we have proved that $U_T = T^{-1}\sum_{t=1}^{T} 1_{[|v_t| \leq d_T]}|x_{ti}||x_{tj}| \xrightarrow{p} 0$ 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 $\hat{Q}_0, \hat{Q}_1, ..., \hat{Q}_G, \hat{\Pi}$ and $\hat{\gamma}$. Since the results $\hat{\Pi} - \Pi_0 = o_p(1)$ and $\hat{\gamma} - \gamma_0 = o_p(1)$ are trivial, we focus on $|\tilde{Q}_0 - Q_0| = o_p(1)$. Let be $Q_{0T} = (2c_{0T}T)^{-1}\sum_{t=1}^{T} 1_{[-c_{0T} \leq v_t \leq c_{0T}]}x_t x_t'$ and $\tilde{Q}_{0T} = (2c_{0T}T)^{-1}\sum_{t=1}^{T} 1_{[-\hat{c}_{0T} \leq \hat{v}_t \leq \hat{c}_{0T}]}x_t x_t'$. 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). One can show easily by using the Mean Value Theorem that $E(Q_{0T}) = E\left\{T^{-1}\sum_{t=1}^{T} f(\xi_T|x_t)x_t x_t'\right\}$,where $-c_{0T} \leq \xi_T \leq c_{0T}$. Noting that $\xi_T = o_p(1)$, it can be shown that $|E(Q_{0T}) - Q_0| = o(1)$ by Minkowski inequality and Assumptions 2(i) and 2(ii). Using a 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 $|\tilde{Q}_{0T} - Q_{0T}|$ is given by $|(2c_{0T}T)^{-1}\sum_{t=1}^{T}(1_{[-\hat{c}_{0T} \leq \hat{v}_t \leq \hat{c}_{0T}]} - 1_{[-c_{0T} \leq v_t \leq c_{0T}]})x_t x_t'| \leq (2c_{0T}T)^{-1}\sum_{t=1}^{T}(1_{[|v_t + c_{0T}| \leq d_T]} + 1_{[|v_t - c_{0T}| \leq d_T]})|x_{ti}||x_{tj}| = U_{1T} + U_{2T}$, 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 the last step, 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 the desired result: $|\hat{Q}_0 - Q_0| = o_p(1)$. The same argument can be applied to show $|\hat{Q}_j - Q_j| = o_p(1)$ for $j = 1, \ldots, G$. Therefore, we have $\hat{D} \xrightarrow{p} D$. *QED.*

C. APPENDIX. SIMULATION DESIGN

The structural system is given by $B\begin{bmatrix} y_t' \\ Y_t' \end{bmatrix} + \Gamma x_t' = U_t'$, where $\begin{bmatrix} y_t' \\ Y_t' \end{bmatrix}$ 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, $B \equiv \begin{bmatrix} 1 & -0.5 \\ -0.7 & 1 \end{bmatrix}$ and $\Gamma \equiv \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & -0.5 \end{bmatrix}$. 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 eq. 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 in a matrix representation $\begin{bmatrix} y & Y \end{bmatrix} B' = -X\Gamma' + U$, which gives the following reduced form equations $\begin{bmatrix} y & Y \end{bmatrix} = X \begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} + \begin{bmatrix} v & V \end{bmatrix}$, where $\begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} \equiv -\Gamma' (B')^{-1}$ and $\begin{bmatrix} v & V \end{bmatrix} \equiv U (B')^{-1}$. Given the specification of B, 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 $\begin{bmatrix} v & V \end{bmatrix}$ in the above reduced form equations are generated in such a way 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 $\mathrm{N}(0, I_T), t(3)$ and LN(0,1) with correlation -0.2, 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 mean $(0.5, 1, -0.1)'$, variances equal to 1, $cov(x_2, x_3) = 0.3, cov(x_2, x_4) = 0.1$ and $cov(x_3, x_4) = 0.2$. Once we obtain X, $\begin{bmatrix} v & V \end{bmatrix}$ and $\begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix}$, we can generate the endogenous variables $\begin{bmatrix} y & Y \end{bmatrix}$ using the reduced-form equations.

The One-Stage Quantile Regression estimator, without correcting for the endogeneity problem, is $\tilde{\alpha} \in \arg\min \sum_{t=1}^{T} \rho_\theta(y_t - [Y_t, x_t]'\alpha)$. The DSQR$(\theta, q)$ is defined by (4). We have chosen 5 values (0.05, 0.25, 0.50, 0.75, 0.95) for θ.

Table 1. Simulation Means and Standard Deviations of the Deviations from the True Value with One Step Quantile Estimator: N(0,1)

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\hat{\gamma}$	Mean	-0.45	-0.43	-0.43	-0.44	-0.44
		Std	0.29	0.18	0.16	0.18	0.29
	$\hat{\beta}_0$	Mean	1.30	1.31	1.38	1.44	1.52
		Std	1.40	0.74	0.55	0.50	0.56
	$\hat{\beta}_1$	Mean	0.17	0.17	0.16	0.16	0.16
		Std	0.35	0.22	0.20	0.22	0.34
$T = 300$	$\hat{\gamma}$	Mean	-0.44	-0.44	-0.44	-0.44	-0.44
		Std	0.11	0.07	0.06	0.07	0.11
	$\hat{\beta}_0$	Mean	1.23	1.32	1.39	1.46	1.57
		Std	0.54	0.28	0.22	0.20	0.22
	$\hat{\beta}_1$	Mean	0.17	0.16	0.16	0.17	0.16
		Std	0.13	0.09	0.08	0.09	0.13

Table 2. Simulation Means and Standard Deviations of the Deviations from the True Value with One Step Quantile Estimator: t (3)

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\hat{\gamma}$	Mean	-0.58	-0.49	-0.48	-0.50	-0.59
		Std	0.44	0.16	0.13	0.16	0.43
	$\hat{\beta}_0$	Mean	1.89	1.53	1.55	1.64	1.89
		Std	2.34	0.65	0.48	0.48	1.23
	$\hat{\beta}_1$	Mean	0.20	0.18	0.18	0.19	0.20
		Std	0.78	0.27	0.22	0.26	0.73
$T = 300$	$\hat{\gamma}$	Mean	-0.57	-0.50	-0.49	-0.50	-0.57
		Std	0.15	0.06	0.05	0.06	0.15
	$\hat{\beta}_0$	Mean	1.92	1.56	1.57	1.64	1.74
		Std	0.84	0.26	0.18	0.19	0.37
	$\hat{\beta}_1$	Mean	0.22	0.18	0.18	0.18	0.21
		Std	0.28	0.10	0.08	0.11	0.27

Table 3. Simulation Means and Standard Deviations of the Deviations from the True Value with One Step Quantile Estimator: LN(0,1)

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\hat{\gamma}$	Mean	0.38	0.47	0.52	0.54	0.56
		Std	0.15	0.09	0.07	0.05	0.05
	$\hat{\beta}_0$	Mean	-0.34	-0.18	0.02	0.29	0.81
		Std	0.18	0.09	0.07	0.10	0.25
	$\hat{\beta}_1$	Mean	-0.14	-0.17	-0.19	-0.20	-0.21
		Std	0.10	0.07	0.07	0.08	0.12
$T = 300$	$\hat{\gamma}$	Mean	0.35	0.48	0.53	0.56	0.58
		Std	0.08	0.04	0.03	0.02	0.02
	$\hat{\beta}_0$	Mean	-0.31	-0.18	0.02	0.31	0.92
		Std	0.09	0.04	0.03	0.04	0.10
	$\hat{\beta}_1$	Mean	-0.13	-0.18	-0.19	-0.21	-0.21
		Std	0.04	0.03	0.03	0.03	0.05

Table 4(a). Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q = 0.1)$: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\hat{\gamma}$	Mean	-0.06	0.01	0.01	0.01	-0.03
		Std	0.80	0.64	0.40	0.48	0.93
	$\hat{\beta}_0$	Mean	-0.07	-0.11	-0.02	0.04	0.37
		Std	2.65	2.08	1.26	1.55	3.05
	$\hat{\beta}_1$	Mean	0.03	0.01	0.00	-0.01	0.03
		Std	0.52	0.30	0.26	0.30	0.50
T = 300	$\hat{\gamma}$	Mean	0.01	0.01	0.01	0.01	0.00
		Std	0.24	0.15	0.14	0.15	0.23
	$\hat{\beta}_0$	Mean	-0.11	-0.03	-0.02	0.00	0.07
		Std	0.80	0.48	0.45	0.48	0.76
	$\hat{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.17	0.10	0.10	0.11	0.16

Table 4(b). Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q = 0.5)$: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\hat{\gamma}$	Mean	-0.07	0.01	0.00	0.01	-0.03
		Std	0.78	0.64	0.40	0.48	0.97
	$\hat{\beta}_0$	Mean	0.15	-0.07	0.00	-0.02	0.14
		Std	2.58	2.10	1.26	1.54	3.18
	$\hat{\beta}_1$	Mean	0.04	0.01	0.00	0.00	0.03
		Std	0.50	0.29	0.26	0.30	0.50
T = 300	$\hat{\gamma}$	Mean	0.01	0.01	0.01	0.01	0.00
		Std	0.25	0.15	0.14	0.15	0.23
	$\hat{\beta}_0$	Mean	-0.04	-0.02	-0.02	-0.02	0.01
		Std	0.80	0.48	0.45	0.48	0.74
	$\hat{\beta}_1$	Mean	-0.01	0.00	0.00	0.00	0.00
		Std	0.17	0.10	0.10	0.11	0.16

Table 4(c). Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q = 1)$: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\hat{\gamma}$	Mean	-0.04	0.02	0.00	0.00	-0.04
		Std	0.85	0.68	0.41	0.49	0.88
	$\hat{\beta}_0$	Mean	0.11	-0.08	-0.01	-0.02	0.11
		Std	2.82	2.22	1.30	1.59	2.87
	$\hat{\beta}_1$	Mean	0.02	0.00	0.00	0.00	0.02
		Std	0.53	0.31	0.26	0.31	0.51
T = 300	$\hat{\gamma}$	Mean	0.01	0.01	0.01	0.00	0.00
		Std	0.25	0.15	0.14	0.15	0.23
	$\hat{\beta}_0$	Mean	-0.03	-0.02	-0.02	-0.02	0.01
		Std	0.81	0.49	0.46	0.50	0.75
	$\hat{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.18	0.10	0.10	0.11	0.16

Table 5. Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q=1)$: t(3).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\hat{\gamma}$	Mean	-0.31	-0.02	0.00	-0.05	-0.32
		Std	1.40	0.73	0.49	0.81	1.35
	$\hat{\beta}_0$	Mean	0.80	0.07	0.00	0.19	1.26
		Std	4.76	2.26	1.56	2.55	4.34
	$\hat{\beta}_1$	Mean	0.14	0.00	0.00	0.03	0.13
		Std	1.15	0.37	0.31	0.44	1.16
T = 300	$\hat{\gamma}$	Mean	-0.06	0.00	-0.01	0.00	-0.07
		Std	0.76	0.19	0.15	0.19	0.63
	$\hat{\beta}_0$	Mean	0.14	0.00	0.04	0.02	0.23
		Std	2.46	0.62	0.49	0.62	2.17
	$\hat{\beta}_1$	Mean	0.03	-0.01	0.00	-0.01	0.02
		Std	0.41	0.13	0.11	0.14	0.41

Table 6. Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q=1)$: LN(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\hat{\gamma}$	Mean	0.00	0.00	0.02	0.12	0.49 (0.47)
		Std	0.08	0.11	0.34	0.70	0.44 (0.34)
	$\hat{\beta}_0$	Mean	-0.06	-0.05	-0.02	0.00	0.37 (0.40)
		Std	0.05	0.07	0.18	0.31	0.78 (0.60)
	$\hat{\beta}_1$	Mean	0.00	0.00	-0.01	-0.04	-0.19 (-0.20)
		Std	0.06	0.08	0.15	0.27	0.49 (0.46)
T = 300	$\hat{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.29 (0.30)
		Std	0.02	0.04	0.07	0.15	0.52 (0.33)
	$\hat{\beta}_0$	Mean	-0.08	-0.06	-0.04	0.01	0.28 (0.32)
		Std	0.02	0.03	0.05	0.10	0.47 (0.43)
	$\hat{\beta}_1$	Mean	0.00	0.00	0.00	0.00	-0.10 (-0.12)
		Std	0.02	0.03	0.05	0.11	0.40 (0.33)

Table 7. Simulation Means and Standard Deviations of the Deviations from the True Value with $DSQR(\theta, q=1)$ and $2SLS$ with a single outlier: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95	$2SLS$
T = 50	$\hat{\gamma}$	Mean	-0.04	0.02	0.01	0.01	0.02	0.24
		Std	0.86	0.64	0.42	0.52	1.56	2.08
	$\hat{\beta}_0$	Mean	0.12	-0.05	0.00	0.02	0.53	0.23
		Std	2.89	2.05	1.33	1.67	4.91	6.51
	$\hat{\beta}_1$	Mean	0.03	0.00	0.00	0.00	0.01	-0.01
		Std	0.53	0.31	0.27	0.33	1.47	1.54
T = 300	$\hat{\gamma}$	Mean	0.01	0.01	0.01	0.00	0.00	0.04
		Std	0.25	0.15	0.14	0.15	0.24	0.32
	$\hat{\beta}_0$	Mean	-0.03	-0.02	-0.02	-0.01	0.03	0.04
		Std	0.81	0.49	0.46	0.50	0.79	1.03
	$\hat{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.01	0.01
		Std	0.18	0.11	0.10	0.11	0.17	0.25