UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/1

TWO-STAGE QUANTILE REGRESSION

by Tae-Hwan Kim and Christophe Muller

January 2000 DP 00/1

Abstract: We present in this paper the asymptotic properties of two-stage quantile regression estimators. These results permit valid inferences in structural models estimated using quantile regressions, in which the possible endogeneity of some explanatory variables is treated via ancilliary predictive equations. Simulation results illustrate the usefulness of this approach.

Résumé: Nous présentons dans ce papier les propriétés asymptotiques des estimateurs en deux étapes pour les régressions quantiles. Ces résultats permettent des inférences valides dans les modèles structurels estimés à partir de régressions quantiles, dans lesquels la possible endogénéité de certaines variables explicatives est traité via des équations prédictives auxilliaires. Des résultats de simulations illustrent l'utilité de cette approche.

Introduction

Quantile regression and least absolute deviation estimators have recently become very popular estimation methods. They are often used in the domains of wage and living standard analysis (Buchinski (1995, 98), Machado and Mata (1997), Muller (1997), Jalan and Ravallion (1998), Disney and Gosling (1998), Anderson and Pomfret (1999), Muller (1999), Nielsen and Rosholm (1999)), but also for analysis of firm data (Mata and Machado (1996), Machado(1997)), as well as in non-economic domains (Lipsitz et al. (1997)).

The popularity of these methods relies on two sets of properties. Firstly, in contrast with usual GMM or MLE methods, they provide robust estimates, particularly for misspecification errors related to heteroscedasticity, non-normality and other error term misspecification, but also for measurement error problems. Secondly, they allow the researcher to concentrate her attention on specific parts of the distribution of interest. This is the case when the distribution of interest is the conditional distribution of the dependent variable. This is also appropriate in living standard analysis when the living standard equations contribute only to a small part of the variance of living standards, so that the distribution of living standard and the distribution of errors are close. Then, when using quantiles in the lower tail of the distribution, the estimated relations would result mostly from the situation of the poor and only marginally from that of the rich. This would avoid mixing heterogeneous economic processes that are specific to different parts of the living standard distribution.

The equations that researchers want to estimate often describe explanatory relations in which some independent variables are endogenous. A typical case is the introduction of socio-economic variables, such as the occupation or the education of the individual, in living standard equations or wage equations. Indeed, these variables are also sometimes predetermined by the living standard of the individual, or these variables and the living standards may be the object of simultaneous choices. Other sources of endogeneity such as measurement errors frequently occur.

In linear models estimated using LS methods, the usual response to such situation is to replace the endogenous explanatory variables by predictions using ancilliary equations based on other exogenous variables and to modify the formula of the asymptotic covariance matrix of estimates (the IV method). The most famous method is based on the two-stage least square estimator whose conditions for consistency and asymptotic normality are known (Malinvaud (1970), Amemiya (1985)). In more complex nonlinear models, other interesting two-stage estimators relying on a first step of predictions for endogenous explanatory variables (sometimes describing selection processes) have been developed and the conditions for their asymptotic properties have been clarified (Heckman (1976), Newey (1985), Newey (1989), Pagan (1986), Newey (1994)).

Amemiya (1982) and Powell (1983) have treated the case of the two-stage least-absolute deviations (2SLAD). The theoretical literature on quantile regression and LAD estimators is extensive since the seminal paper by Koenker and Basset. Several authors have studied the asymptotic behaviour of these estimators (Koenker and Basset (1978, 82), Basset and Koenker (1978, 86), Powell (1983), Weiss (1990), Phillips (1991), Pollard (1991)). However, apart from the attempts of Amemiya and Powell in the 2SLAD case, two-stage estimations of quantile regressions (2SQR) have not been investigated up to now footnote .

We present in this paper the asymptotic properties of two-stage quantile regression estimators. Section 2 discusses the model and the assumptions. In section 3 we derive the asymptotic representation of the estimators. We prove in section 4 the asymptotic normality with quantile regression predictions and discuss the asymptotic variance matrix. We discuss the asymptotic bias in section 5. We analyse in section 6 the the asymptotic normality and the asymptotic covariance matrix with LS predictions. We present simulation results in section 7. Finally, section 8 concludes.

The Model

Let us suppose that we are interested in the structural parameter (α_0) in an equation that is given in the following matrix form for a sample of T observations:

$$\begin{aligned} y &= Y\gamma_0 + X_1\beta_0 + u \\ &\equiv Z\alpha_0 + u \end{aligned} \qquad \#$$

where $[y, Y]$ is a $T \times (G+1)$ matrix of endogenous variables, X_1 is $T \times K_1$ matrix of exogenous variables, $Z \equiv [Y, X_1]$, $\alpha_0' \equiv [\gamma_0', \beta_0']$, and u is a $T \times 1$ vector. We denote by X_2 the matrix of $K_2 (\equiv K - K_1)$ exogenous variables absent from the first equation. These notations and conventions correspond to those in Powell (1983) for the case of two-stage LAD estimation. Let us assume that Y has a reduced-form representation:

$$Y = X\Pi_0 + V \qquad \#$$

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. We specify now the data generating process.

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

Then, using eqs. ref: eq0 and ref: eq1 , y also has a reduced form representation:

$$y = X\pi_0 + v \qquad \#$$

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 $v \equiv u + V\gamma_0$. Equations ref: eq1 and ref: eq2 are the basis of the first stage estimation that yields some estimators $\hat{\pi}$, $\hat{\Pi}$ respectively of π_0, Π_0. We impose the following assumptions on the first step estimators.

Assumption 2. $T^{1/2}(\hat{\pi}-\pi_0) = O_p(1)$ *and* $T^{1/2}(\hat{\Pi}-\Pi_0) = O_p(1)$.

Following the standard literature on quantile regressions, we 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 $\psi_\theta(z) \equiv \theta - 1_{[z\leq 0]}$ where $1_{[.]}$ is the Kronecker index. As a natural extension of Amemiya (1982) and Powell (1983), we define the Two-Stage Quantile Regression (2SQR(θ, q)) estimator $\hat{\alpha}$ of α_0 as a solution to the following minimisation programme.

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

where y_t and X_t' are the t^{th} elements in y and X respectively and q is a strictly positive constant chosen in advance by the researcher.

In the next section we discuss the asymptotic representation of the 2SQR(θ, q) estimator $\hat{\alpha}$. We shall show that the following conditions are sufficient for the asymptotic representation.

Assumption 3. $\max_{t,k} \; T^{-1/2}|X_{tk}| \to 0$ *and* $T^{-1}\sum_{t=1}^{T} X_t X_t' \to Q$ *where Q is finite and positive-definite*.

Assumption 4. $H(\Pi_0)$ *is full column rank*.

Assumption 5. v_t *has a continuous density* $f(\lambda)$ *and* $f(0) > 0$.

Assumption 6. $E(\psi_\theta(v_t)) = 0$ *or equivalently* $\int_{-\infty}^{0} f(\lambda)d\lambda = \theta$.

Assumptions 3-5 are standard in the literature. Assumption 6 is a generalization of Powell’s assumption for two-stage LAD estimation that zero is the median of the distribution of v_t. When equation ref: eq2 is estimated using the θ^{th} quantile regression, Assumption 6 is necessary for the asymptotic normality of the first stage estimation for the dependent variable y.

When there is a constant term in the model, Assumption 6 is a restriction that can be considered as an identification condition on the coefficient of the constant. Indeed, models such that $E\psi_\theta(v) = 0$ and $E\psi_\theta(v) \neq 0$ correspond to isomorphic statistical structures that distinguish themselves only by the value of the constant term. They are observationally equivalent structures. Therefore, it is possible to impose $E\psi_\theta(v) = 0$, and thus to fix the value of the constant, without loss of generality.

The Asymptotic Representation

The first step of the analysis is the derivation of an asymptotic representation of the

2SQR(θ, q). For this purpose, we consider the following data generating process that is deduced from equation ref: eq2 .

$$\tilde{y}_t = \tilde{X}'_t \alpha_0 + \tilde{\epsilon}_t \quad \#$$

where $\tilde{y}_t \equiv q y_t + (1-q) X'_t \pi_0$ and $\tilde{X}'_t \equiv X'_t H(\Pi_0)$.

Simple algebra shows that $\tilde{\epsilon}_t = q v_t$ and $E(\psi_\theta(\tilde{\epsilon}_t)) = 0$.

The quantile estimation of α_0 from equation ref: eq4 is not possible because both the dependent and independent variables include unknown auxiliary parameters (π_0, Π_0). Nevertheless, equation ref: eq4 provides us with insight on the estimation of α_0 when the true auxiliary parameters (π_0, Π_0) are known. Most asymptotic results for quantile regressions will still be valid with slight modifications when replacing π_0, Π_0 with their consistent and asymptotically normal estimator $\hat{\pi}$, $\hat{\Pi}$. Equation ref: eq4 is the basis of the derivation of the asymptotic representation because it allows the direct application of Bickel's (1975) results. Indeed, it is easy to see that all the conditions for Lemma 4.1 in Bickel (1975) are satisfied: (i) $\tilde{\epsilon}_t$ is i.i.d. (by Assumption 1), (ii) $T^{-1}\sum_{t=1}^{T} \tilde{X}_t \tilde{X}'_t \to Q_{zz} \equiv H(\Pi_0)' Q H(\Pi_0)$ (by Assumption 3) and Q_{zz} is positive-definite (by Assumptions 3-4), (iii) $\max_{t,k} T^{-1/2} | \tilde{X}_{tk} | \to 0$ (by assumption 3), (iv) $\psi_\theta(z - T^{-1/2} \tilde{X}'_t \Delta)$-$\psi_\theta(z)$ satisfies Condition C1 in Bickel (1975).

Next, we define

$$M_T(\Delta) \equiv T^{-1/2} \sum_{t=1}^{T} \tilde{X}_t \psi_\theta(\tilde{\epsilon}_t - T^{-1/2} \tilde{X}'_t \Delta)$$

where Δ is a $(G + K_1) \times 1$ vector. A direct application of Bickel's lemma yields the following lemma.

lemma

Lemma ref: lem1 is slightly different from Lemma 4.1 in Bickel (1975) because we replace $E(M_T(\Delta) - M_T(0))$ with its limit $-\omega Q_{zz}\Delta$. The following lemma is an extension of Lemma ref: lem1 which will be used for the asymptotic representation of the 2SQR(θ, q).

lemma

Here, $\hat{\triangle}_2$ is a random variable instead of a fixed vector like $\triangle$ in Lemma ref: lem1 . We combine Lemmas ref: lem1 - ref: lem2 and Assumption 2 to obtain the following asymptotic representation for the the 2SQR(θ, q). All proofs are provided in Appendix 1.

proposition

The asymptotic representation shows that the asymptotic distribution of the second stage estimator $T^{1/2}(\hat{\alpha} - \alpha_0)$ depends on the asymptotic distribution of the first stage estimators $T^{1/2}(\hat{\pi} - \pi_0)$ and $T^{1/2}(\hat{\Pi} - \Pi_0)\gamma_0$ footnote . Naturally, if $q = 1$, the influence of $\hat{\pi}$ disappears. The asymptotic representation of the 2SQR estimator is composed of three additive terms. The first term does not perturb consistency under Assumption 6 and represent the contribution of the second stage to the uncertainty of the estimator. The second and third terms represent the contributions of respectively $\hat{\pi}$ and $\hat{\Pi}$ to the uncertainty of the estimator.

The asymptotic representation can be easily extended to a more general case where the first

step estimators $\hat{\pi}$ and $\hat{\Pi}$ are asymptotically biased. This situation arises for example when the LS estimation method is used in the first step. For this purpose we define the following assumption.

[noindent] **Assumption 2$'$**. There exist $|B_\pi| < \infty$ and $|B_\Pi| < \infty$ such that $T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) = O_p(1)$ and $T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi) = O_p(1)$.

[noindent] Using Lemmas ref: lem1 - ref: lem2 and Assumption $2'$, we obtain the following asymptotic representation with a possible bias.

proposition

[noindent] Proof: See Appendix 1.

Asymptotic Normality and Covariance Matrix with Quantile-Regression Predictions

To obtain the asymptotic representations of the 2SQR(θ, q) in Section 3, it was sufficient to

impose Assumption 2 or $2'$ regarding the preliminary estimators $\hat{\pi}$ and $\hat{\Pi}$. We now need to specify an estimation procedure to obtain the asymptotic representations of $T^{1/2}(\hat{\pi} - \pi_0)$ and $T^{1/2}(\hat{\Pi} - \Pi_0)$, which will be substituted into the asymptotic representation of $T^{1/2}(\hat{\alpha} - \alpha_0)$. Powell (1983) used both LS and LAD methods to compute preliminary estimators of two-stage LAD estimator.

In our case, however, using the LS estimation for π_0 and Π_0 in the first step yields asymptotic normality only for the slope coefficients of the 2SQR(θ, q) estimator $\hat{\alpha}$. As a matter of fact, $T^{1/2}(\hat{\alpha} - \alpha_0)$ diverges to infinity owing to an asymptotic bias on the intercept coefficient induced by the nonvanishing difference between quantile and mean, $E(\psi_\theta(v_t)) - E(v_t)$. If $\theta = 1/2$ and the distribution is symmetric, then the bias vanishes, as in Powell (1983). Also, using the quantile regression method to estimate π_0 and Π_0 will cause the bias to disappear. Hence, an approach to the treatment of endogeneity in quantile regressions is to use quantile regressions based *on the same quantile* for generating exogenous predictions of the endogenous independent variables. For the quantile regressions in the first step, we impose some additional assumptions.

Assumption 5$'$. *V_{jt} has a continuous density $g_j(\lambda)$ and $g_j(0) > 0$* for $j = 1, 2, ..., G$.

Assumption 6$'$. $E(\psi_\theta(V_{jt})) = 0$ *or equivalently* $\int_{-\infty}^{0} g_j(\lambda) d\lambda = \theta$ *for* $j = 1, 2, ..., G$.

lemma

(1) Suppose that Assumptions 1,3,5 and 6 hold. Then,

$$T^{1/2}(\hat{\pi} - \pi_0) = Q^{-1}T^{-1/2}\sum_{t=1}^{T} X_t f(0)^{-1}\psi_\theta(v_t) + o_p(1).$$

(2) Suppose that Assumptions 1,3,5$'$ and 6$'$ hold. Then,

$$T^{1/2}(\hat{\Pi}_j - \Pi_{0j}) = Q^{-1}T^{-1/2}\sum_{t=1}^{T} X_t g_j(0)^{-1}\psi_\theta(V_{jt}) + o_p(1).$$

Proof: Direct consequence of Proposition ref: arbase applied with $q = 1$ and no independent endogenous variables. QED.

We now present the main theorem of the asymptotic normality of the 2SQR(θ, q). We apply the Liapounov's CLT, for which we need the following additional assumptions.

Assumption 7. *There exists a positive constant* $\triangle_3$ *such that* $\|X_t\| \le \triangle_3 < \infty$ *for all t.*

Assumption 8. *There exists positive constants* δ *and* Δ_4 *such that* $0 < E|\eta_t|^{2+\delta} < \Delta_4 < \infty$ *where* $\eta_t \equiv f(0)^{-1}\psi_\theta(v_t) - \xi_t$
and $\xi_t \equiv [g_1(0)^{-1}\psi_\theta(V_{1t}), \ldots, g_G(0)^{-1}\psi_\theta(V_{Gt})]'\gamma_0$.

proposition
Suppose that Assumptions 1, 3-6,5$'$-6$'$, *and* 7-8 *hold. Then,* $T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, C)$ *where* $C \equiv \sigma_0^2 Q_{zz}^{-1}$ *and* $\sigma_0^2 \equiv E(\eta_t^2)$.

Proof: See Appendix 1.

We now propose a consistent estimator for the asymptotic covariance matrix C so as to be able to test a hypothesis on the structural parameter α_0. We first discuss the estimation of the densities of v_t and V_{jt} at zero, denoted by $f(0)$ and $g_j(0)$ respectively. As shown in Powell (1984, 1985), the estimators

$$\hat{f}(0) \equiv \frac{1}{2\ _T}\sum_{t=1}^{T} 1_{[-\ _T \le \hat{v}_t \le\ _T]}$$

$$_j(0) \equiv \frac{1}{2\ _{jT}}\sum_{t=1}^{T} 1_{[-\ _{jT} \le \hat{V}_{jt} \le\ _{jT}]}$$

are consistent for $f(0)$ and $g_j(0)$. The following assumptions are sufficient for the consistency.

Assumption 9. (1) *There exists a stochastic sequence* $\{\ _T\}$ (*depending on the data*) *and a non-stochastic sequence* $\{c_T\}$ *such that*:

(*i*) $\frac{\ _T}{c_T} \xrightarrow{p} 1$ (*ii*) $c_T = o_p(1)$ (*iii*) $c_T^{-1} = o_p(T^{1/2})$.

(2) *There exist a stochastic sequence* $\{\ _{jT}\}$ (*depending on the data*) *and a non-stochastic sequence* $\{c_{jT}\}$ *for each* $j = 1, 2, ..., G$ *such that*:

(*i*) $\frac{_{jT}}{c_{jT}} \xrightarrow{p} 1$ (*ii*) $c_{jT} = o_p(1)$ (*iii*) $c_{jT}^{-1} = o_p(T^{1/2})$.

Assumption 10. *Density restrictions*

(*i*) *There exists* f_0 *such that* $f(\lambda) \leq f_0$ *for all* λ.

(*ii*) *For each* $j = 1, 2, ..., G$, *there exists* g_{0j} *such that* $g_j(\lambda) \leq g_{0j}$ *for all* λ.

Examples of sequences $_T$ and $_{jT}$ are provided in Powell (1984). Conditions (i)-(ii) in Assumption 10 are typical regularity conditions for densities, used in the quantile literature.

Next, we focus on the estimation of σ_0^2. Using the binomial structure incorporated in the check function (i.e. $E(\psi_\theta(v_t)^2) = \theta(1-\theta)$), σ_0^2 can be expressed as:

$$\sigma_0^2 = \frac{\theta(1-\theta)}{f(0)^2} + \gamma_0^{'}\Omega\gamma_0 - \frac{2}{f(0)}\gamma_0^{'}\Gamma$$

where the matrix Ω has a typical element

$\Omega_{ij} \equiv g_i(0)^{-2}\theta(1-\theta)1_{[i=j]} + g_i(0)^{-1}g_j(0)^{-1}(\delta_{ij} - \theta^2)1_{[i\neq j]}$ and

$\delta_{ij} \equiv E(1_{[V_{it}\leq 0]}1_{[V_{jt}\leq 0]})$ and the vector Γ has a typical element $\Gamma_i \equiv g_i(0)^{-1}(\delta_i - \theta^2)$ with $\delta_i \equiv E(1_{[v_t\leq 0]}1_{[V_{it}\leq 0]})$. Using the plug-in method, the proposed estimator is:

$$\hat{\sigma}^2 \equiv \frac{\theta(1-\theta)}{\hat{f}(0)^2} + \hat{\gamma}^{'}\hat{\Omega}\hat{\gamma} - \frac{2}{\hat{f}(0)}\hat{\gamma}^{'}\hat{\Gamma}$$

where the matrix $\hat{\Omega}$ has a typical element

$\hat{\Omega}_{ij} \equiv {}_i(0)^{-2}\theta(1-\theta)1_{[i=j]} + {}_i(0)^{-1}\ {}_j(0)^{-1}(\hat{\delta}_{ij} - \theta^2)1_{[i\neq j]}$,

$\hat{\delta}_{ij} \equiv T^{-1}\sum_{t=1}^{T} 1_{[\hat{V}_{it}\leq 0]}1_{[\hat{V}_{jt}\leq 0]}$ and $\hat{V}_{jt} \equiv Y_{jt} - X_t^{'}\hat{\Pi}_j$ and the vector $\hat{\Gamma}$ has a typical element $\hat{\Gamma}_i \equiv {}_i(0)^{-1}(\hat{\delta}_i - \theta^2)$, $\hat{\delta}_i \equiv T^{-1}\sum_{t=1}^{T} 1_{[\hat{v}_t\leq 0]}1_{[\hat{V}_{it}\leq 0]}$ and $\hat{v}_t \equiv y_t - X_t^{'}\hat{\pi}$. Note that directly substituting consistent estimators in the formula of σ_0^2 in Proposition ref: normqr is likely to yield a less efficient estimator.

lemma

Proof: See Appendix 1.

The covariance matrix can be consistently estimated using the following estimator:

$$\equiv \hat{\sigma}^2\hat{Q}_{zz}^{-1}$$

where $\hat{Q}_{zz} \equiv T^{-1}\sum_{t=1}^{T} H(\hat{\Pi})X_tX_t^{\prime}H(\hat{\Pi})^{\prime}$. The consistency of $\hat{Q}_{zz}$ is straightforward by Assumption 3, which together with Lemma ref: cvsigma delivers the desired result: $\xrightarrow{p} C$.

Asymptotic Bias

We start from the following assumption that is satisfied for example by least-square estimators for the first step of the estimation under Assumption 6 (and of course if alternatively $Ev_t = 0$), or by quantile regression estimators under the alternative assumption that $Ev_t = 0$.

Assumption H: *The first-step estimators of the slope coefficients are consistent*.

According to the asymptotic representation, a bias in $\hat{\pi}$ and is $\hat{\Pi}$ transmitted to the 2SQR estimator through the matrix RQ where

$R = [H(\Pi_0)'QH(\Pi_0)]^{-1}H(\Pi_0)'$.

To simplify the analysis, we rewrite eq. ref: eq0 so as to put the constant term in first position. This yields

$$\begin{aligned} y &= X_1\beta_0 + Y\gamma_0 + +u \\ &\equiv Z\alpha_0 + u \end{aligned} \qquad \#$$

with a new ordering for the components of Z and α_0. Then, the matrix $H(\Pi_0)$ still defined by $\pi_0 \equiv H(\Pi_0)\alpha_0$ is equal to

$$H(\Pi_0) \equiv \begin{bmatrix} I_{K_1} & \Pi_0 \\ 0_{K_2 \times K_1} & \end{bmatrix}.$$

To separate the possible bias on the intercept of the first-stage estimators, we decompose both matrix Q and the first step estimators. Then, $Q \equiv \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ where Q_1 is a $K \times 1$ matrix and Q_2 is a $K \times (K-1)$ matrix, and for example for $\hat{\pi} - \pi_0 \equiv \begin{bmatrix} \hat{\pi}_{(1)} - \pi_{0(1)} \\ \hat{\pi}_{(2)} - \pi_{0(2)} \end{bmatrix}$, where $\hat{\pi}_{(1)}$ is the estimator of the constant coefficient.

Then, $RQ(\hat{\pi} - \pi_0) = RQ_1(\hat{\pi}_{(1)} - \pi_{0(1)}) + RQ(\hat{\pi}_{(2)} - \pi_{0(2)})$ where the second term in the right-hand-side term is asymptotically unbiased by hypothesis. It is therefore necessary and sufficient to study the product RQ_1 to understand the generation of a possible asymptotic bias of $\hat{\alpha}$. Because of the presence of a constant term, vector Q_1 can be further decomposed and characterised by the equality of their first coordinate to one. By definition:

$$Q_1 \equiv \begin{bmatrix} 1 \\ Q_{21} \end{bmatrix}.$$

The formal calculation footnote of the product RQ_1 using product operation by block and inverse operation by block yields a very simple expression owing to the presence of 1 in Q_1 and to the special form of $H(\Pi_0)$, although intermediate matrices appearing in the calculus correspond to complicated and long expressions. We obtain the following remarkable result.

$$RQ_1 = \begin{bmatrix} 1 \\ 0_{(K_1+G-1)\times 1} \end{bmatrix} \qquad \#$$

Eq. ref: eqr1 implies that the only coordinate of $\hat{\alpha}$ for which there is a possible asymptotic bias corresponds to the intercept. Moreover, this asymptotic bias is equal to $(1-q)$ times the asymptotic bias in the intercept in $\hat{\pi}$ minus the asymptotic bias in $\hat{\Pi}\gamma_0$.

Several favourable situations may occur. First, empirical economists are generally interested in the slope components of $\hat{\alpha}$ rather than in its intercept coefficient. Then, any first-step estimation method satisfying our mentioned assumptions will deliver the consistency and the asymptotic normality of the slope coefficients.

Second, in cases where $\hat{\Pi}.\gamma_0$ is not asymptotically biased, for example because the first-step estimation method for Π_0 is such that $T^{1/2}(\hat{\Pi} - \Pi_0)$ is $O_P(1)$, the asymptotic bias of the coefficient of the intercept in $\hat{\alpha}$ is $(1-q)$ times the asymptotic bias of $\hat{\pi}$. Choosing $q = 1$ guaranties that this bias disappears.

Asymptotic Normality and Covariance Matrix with LS Prediction

Complete parameter

In this section, we investigate the use of LS estimation for π_0 and Π_0 in the first step. As mentioned earlier, the problem is that the errors V_t and v_t in eqs. ref: eq1 and ref: eq2 do not have zero expectation under Assumption 6. When the expectation of the regression error is not zero, the normalised LS estimator $T^{1/2}(\tilde{\alpha} - \alpha_0)$ diverges to infinity in general.

First, we define $V_t^* \equiv V_t - E(V_t)$ and $v_t^* \equiv v_t - E(v_t)$. Then, the reduced forms for Y_t and y_t in eqs. ref: eq1 and ref: eq2 can be expressed as

$$Y_t = X_t'\Pi_0^* + V_t^* \qquad \#$$

where $\Pi_0^* \equiv \Pi_0 + B_\Pi$ and $B_\Pi \equiv [E(V_t)', 0', ..., 0']'_{(K\times G)}$

$$y_t = X_t'\pi_0^* + v_t^* \qquad \#$$

where $\pi_0^* \equiv \pi_0 + B_\pi$ and $B_\pi \equiv [E(v_t), 0, ..., 0]'_{(K\times 1)}$. By construction, $E(V_t^*) = E(v_t^*) = 0$. Let $\tilde{\Pi}$ and $\tilde{\pi}$ be the LS estimators based on eqs. ref: redform1 and ref: redform2 respectively. Let $\tilde{\alpha}$ be the 2SQR(θ, q) estimator based on the LS estimators $\tilde{\Pi}$ and $\tilde{\pi}$ in the first step.

It can be shown using Proposition ref: arbias that the asymptotic representation for the 2SQR estimator based on LS predictions is

$$T^{1/2}(\tilde{\alpha} - \alpha_0 - B_\alpha) = Q_{zz}^{-1}H(\Pi_0)'\{T^{-1/2}\sum_{t=1}^{T} X_t q f(0)^{-1}\psi_\theta(v_t)$$
$$+ (1-q)QT^{1/2}(\tilde{\pi} - \pi_0 - B_\pi) - QT^{1/2}(\tilde{\Pi} - \Pi_0 - B_\Pi)\gamma_0\} + o_p(1)$$

where $B_\alpha \equiv Q_{zz}^{-1}H(\Pi_0)'Q\{(1-q)B_\pi - B_\Pi\gamma_0\}$.

Then, the asymptotic normality of $\tilde{\alpha} - \alpha_0 - B_\alpha$ can be easily derived. To obtain it we impose the following assumptions.

Assumption 6$''$. $E(u_t) = 0$.

Assumption 8$'$. *There exists constants* $\delta > 0$ *and* $\Delta_5 > 0$ *such that* $0 < E|\zeta_t|^{\delta+2} < \Delta_5 < \infty$ *where* $\zeta_t \equiv qf(0)^{-1}\psi_\theta(v_t) + u_t - q(v_t - E(v_t))$.

Then, we have

proposition

Suppose that Assumptions 1, 3-6, 6$''$, 7 *and* 8$'$ *hold. Then,*

$$T^{1/2}(\tilde{\alpha} - \alpha_0 - B_\alpha) \xrightarrow{d} N(0, \sigma_0^2 Q_{zz}^{-1})$$

where $\sigma_0^2 \equiv E(\zeta_t^2)$.

Proof: Similar to the proof of Proposition ref: normqr .

Slope parameter

However, since the asymptotic bias affects only the intercept coefficient, it is useful to investigate separately the asymptotic properties of the slope coefficient.

Since we are interested in the slope coefficient, we decompose $\tilde{\Pi} = \begin{bmatrix} \tilde{\Pi}_{(1)} \\ \tilde{\Pi}_{(2)} \end{bmatrix}$ where $\tilde{\Pi}_{(1)}$ is the first $1 \times G$ row and $\tilde{\Pi}_{(2)}$ is the remaining $(K-1) \times G$ matrix and $\tilde{\pi} = \begin{bmatrix} \tilde{\pi}_{(1)} \\ \tilde{\pi}_{(2)} \end{bmatrix}$ where $\tilde{\pi}_{(1)}$ is the first element and $\tilde{\pi}_{(2)}$ is the remaining $(K-1) \times 1$ vector. Hence, $\tilde{\Pi}_{(2)}$ and $\tilde{\pi}_{(2)}$ contain only the slope coefficients. We also decompose $Q^{-1} = \begin{bmatrix} \bar{Q}_1 \\ \bar{Q}_2 \end{bmatrix}$ where $\bar{Q}_1$ is the first $1 \times K$ row and $\bar{Q}_2$ is the remaining $(K-1) \times K$ matrix. Then, it is straightforward to see that

$$T^{1/2}(\tilde{\Pi}_{(2)} - \Pi_{0(2)}) = \bar{Q}_2 T^{-1/2} \sum_{t=1}^{T} X_t V_t^* + o_p(1), \qquad \#$$

$$T^{1/2}(\tilde{\pi}_{(2)} - \pi_{0(2)}) = \bar{Q}_2 T^{-1/2} \sum_{t=1}^{T} X_t v_t^* + o_p(1).$$

where $\Pi_{0(2)}$ and $\pi_{0(2)}$ are the corresponding components of Π_0 and π_0 respectively.

We decompose $\tilde{\alpha} = \begin{bmatrix} \tilde{\alpha}_{(1)} \\ \tilde{\alpha}_{(2)} \end{bmatrix}$ where $\tilde{\alpha}_{(1)}$ is the first element and $\tilde{\alpha}_{(2)}$ is the remaining $(K_1 + G - 1) \times 1$ vector and $\alpha_0 = \begin{bmatrix} \alpha_{0(1)} \\ \alpha_{0(2)} \end{bmatrix}$ likewise. Using eq. ref: asymrepols and the asymptotic representation in Proposition ref: arbias , it can be shown that:

lemma

From this lemma, the asymptotic normality of the slope coefficients is easily derived.

proposition

Suppose that Assumptions 1, 3-6, 6′, 7 *and* 8′ *hold. Then,*

$$T^{1/2}(\tilde{\alpha}_{(2)} - \alpha_{0(2)}) \xrightarrow{d} N(0, \sigma_0^2 R_2 Q R_2').$$

Proof: Similar to the proof of Proposition ref: normqr .

There are some technical difficulties for obtaining a consistent covariance matrix for $\sigma_0^2 R_2 Q R_2'$. First, R_2 is a submatrix of R which can be expressed as

$$R = [H(\Pi_0^* - B_\Pi)' QH(\Pi_0^* - B_\Pi)]^{-1} H(\Pi_0^* - B_\Pi)'.$$

Since the LS estimator $\tilde{\Pi}$ is consistent for Π_0^*, a consistent estimator for R can be obtained only when we have some consistent estimator for the bias term $E(V_t)$ in B_Π. Second, we note that ζ_t can be expressed as

$$\zeta_t = qf(0)^{-1}\psi_\theta(v_t^* + E(v_t)) + u_t - q(v_t^*)$$

Therefore, obtaining a consistent estimator for σ_0^2 requires some consistent estimator for the bias term $E(v_t)$. This can be made by estimating the first-step predictions both using LS and quantile regressions (with the same θ), and using the residuals of these estimations similarly to section 4.

Optimal q

When the LS method is chosen for the first stage, it is possible to propose an optimal choice of q, so as to minimise the variance of the slope coefficients estimator. This optimal choice corresponds to the solution of the following programme.

$$\min_q \quad E(\zeta_t)^2 \text{ subject to } 0 < q \qquad \#$$

The calculation gives the following result.

proposition
Then, the optimal choice of q, q^*, *is*:

a) *If* $E[\psi_\theta(v_t)u_t] \geq f(0)E[(v_t - Ev_t)u_t]$, *then* $q^* = 0^+$,

b) *Otherwise* $q^* = \frac{E[(v_t - Ev_t)u_t] - E\left[f(0)^{-1}\psi_\theta(v_t)u_t\right]}{E\left[f(0)^{-1}\psi_\theta(v_t) - (v_t - Ev_t)\right]^2}$.

Then, information about characteristics of the error terms may enable the analyst to choose a value of q that will enhance the accuracy of the estimates. In the case b) this value can be estimated using empirical analogues as for the estimation of the variance-covariance matrix.

Monte Carlo Simulations

We conduct simulation experiments to investigate the finite sample properties of the 2SQR(θ, q) estimator. The main two focus of the simulation are firstly to compare 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 using our procedure; secondly to assess the size of the asymptotic bias when it exists.

The data generating process used in the simulation study corresponds a simple simultaneous structural model and is shown in Appendix 2. It incorporates two endogenous variables and three exogenous variables including a constant. Here $\beta_0 = (\beta_{00}, \beta_{10})'$.

The performance of the one-stage quantile regression estimator is displayed in Table 1. This estimator is systematically biased in finite samples. The slope parameters (γ_0, β_{10}) are underestimated while the intercept parameter (β_{00}) is overestimated. The results vary little when increasing the number of observations from 50 to 300, although the sampling distribution is more concentrated about the mean.

The results for the 2SQR(θ, q) estimator, denoted ($\hat{\gamma}$, $\hat{\beta}$), based on the quantile prediction in the first step are provided in Table 2. The means of ($\hat{\gamma}, \hat{\beta}$) are much closer to the true parameters than the one-step quantile estimator over all values of θ, although the corresponding standard deviations are generally greater. This implies a trade-off between bias and variance in the choice of the estimator. When $T = 50$ and 100, the 2SQR(θ, q) estimator is biased owing to the small sample size (β_{00} and β_{10} are generally underestimated and γ_0 is generally overestimated). The bias disappears and the standard deviation becomes smaller by half when we increase the sample size from 100 to 300. It is also evident from the table that since the standard deviations are symmetric quadratic functions of θ about $\theta = 0.5$: they become larger as θ becomes closer to 0 or 1. Identical results are obtained for all values of q because the asymptotic representation of $\hat{\alpha}$ does not depend on q when the first stage estimator are derived from quantile regression using the same quantile.

We report the finite sample results for the 2SQR(θ, q) based on the LS predictions in Tables 3-5. Here, the sampling distribution of the 2SQR(θ, q) estimator depends on the value of q and we show the simulation results for q= 0.1, 0.5 and 1. The 2SQR(θ, q) estimator based on LS predictions provides accurate estimates for the slope parameters (γ_0 and β_{10}) over the all values of q and θ. The means are closer to the true slope parameters than for the 2SQR(θ, q) estimator based on quantile predictions, and the standard deviations are smaller. However, the estimates of the intercept parameter (β_{00}) are systematically and substantially biased. When $q = 0.1$, the direction of the bias is inversely related with θ: overestimation for $\theta < 0.5$, no bias for $\theta = 0.5$ and underestimation for $\theta > 0.5$. The opposite phenomenon occurs for $q = 1$ in Table 5. The intercept parameter is not biased for $q = 0.5$ as shown in Table 4 and by Powell (1983).

The 2SQR(θ, q) estimator based on quantile predictions has good finite sample properties, although a too small sample size (50 observations) can seriously degrade its performance. The 2SQR(θ, q) estimator based on the LS prediction is potentially better when we are interested in only slope parameters in the structural equation, or when the sample size is very small, but its intercept coefficient is substantially biased.

Conclusion

We present in this paper the asymptotic properties of two-stage quantile regression estimators.

These results permit valid inferences in structural models estimated using quantile regressions, in which the possible endogeneity of some explanatory variables is treated via ancillary predictive equations.

Amemiya Amemiya, T. (1982), "Two Stage Least Absolute Deviation Estimators," *Econometrica*, 50, 689-711.

Amemiya (2) Amemiya, T. (1985), "The Nonlinear Tow-Stage Least-Squares Estimation," *Journal of Econometrics*, 2, 105-110.

Anderson Anderson, K. and R. Pomfret (1999), "Living Standards during Transition to a Market Economy: The Kyrgyz Republic in 1993 and 1996," Working Paper 99-3, University of Adelaide.

Barrodale Barrodale, I. and F.D.K. Roberts (1974), "Algorithm 478: Solution of an Overdetermined System of Equations in the L_1 Norm," *Communications of the Association for Computing Machinery*, 17, 319-320.

Basset Basset, G.W. and R.W. Koenker (1986) "Strong Consitency of regression Quantiles and related Empirical Processes," *Econometric Theory*, 2, 191-201.

Basset (2) Basset, G.W. and R.W. Koenker (1978) "Asymptotic Theory of Least Absolute

Error Regression," *Journal of American Statistical Association*, 73, 618-622.

Bickel Bickel, P.J. (1975), "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, Vol. 70, N. 350, 428-433.

Buchinski Buchinski, M. (1995), "Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987," *Journal of Econometrics*, 65, 109-154.

Buchinski (2) Buchinski, M. (1998), "Recent Advances in quantile Regression Models. A Practical Guideline for Empirical Research," *The Journal of Human Resources*, XXXIII, 1, 88-126.

Disney Disney, R. and A. Gosling (1998), "Does it Pay to Work in the Public Sector," *Fiscal Studies*, 19 (4), 347-74.

Heckman Heckman, J.J. (1976), "The Common Structure of Statistical Model of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for such Models," *Annals of Economic and Social Measurement*, 5, 475-492.

Jalan Jalan, J. and M. Ravallion (1998), "Transient Poverty in Postreform Rural China," *Joural of Comparative Economics*, 26, 338-357.

Jureckova Jureckova, J. (1977), "Asymptotic Relations of M-Estimates and R-Estimates in Linear Regression Model," *The Annals of Statistics*, 5, 464-472.

Koenker, R. and G. Bassett (1) Koenker, R. and G. Bassett (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Koenker, R. and G. Bassett (2) Koenker, R. and G. Bassett (1982), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43-61.

Lipsitz 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, No. 4, 463-476.

Machado Machado, A.F. (1997), "Box-Cox Quantile Regression and the Distribution of Firm Sizes," mimeo Universidade Nova de Lisboa, February.

Machado-mata Machado, A.F.,and J. Mata (1997), "Earning Functions in Portugal 1982-1994: Evidence from Quantile Regressions," mimeo Universidade Nova de Lisboa, December.

Malinvaud Malinvaud, E. (1970), "The consistency of Nonlinear regressions" *Annals of Mathematical Statistics*, 41, 956-969.

Mata-machado Mata, J. and J.A.F. Machado (1996), "Firm Start-Up Size: A Conditional Quantile Approach," *European Economic Review*, 40, 1305-1323.

Muller Muller, C. (1997), Transient Seasonal and Chronic Poverty of the Peasants: Evidence From Rwanda", CSAE Working paper, University of Oxford.

Muller (2) Muller, C. (1999), "Censored Quantile Regression of Poverty in Rwanda", CREDIT Working Paper, University of Nottingham.

Newey2 Newey, W.K. (1985), "Semiparametric Estimation of Limited Dependent Variables Models with Endogenous Explanatory Variables," *Annales de l'INSEE*, n. 59/60, 219-235.

Newey Newey, W.K. (1989), "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14, 201-206.

Newey3 Newey, W.K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, Vol. 62, No. 6, 1349-1382, November.

Nielsen Nielsen, H.S. and M. Rosholm (1999), "The Effect of Privatization and Market Liberalization on Wage Formation in Zambia in the 1990s: A Quantile Regression Approach," mimeo Aarhus school of Business, Department of Economics.

Pagan Pagan, A.R. (1986), "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies*, 57, 517-538.

Phillips Phillips, P.C.B. (1991), "A Shortcut to LAD Estimator Asymptotics," *Econometric Theory*, Vol. 7, 450-463.

Pollard Pollard, D. (1991), "Asymptotic for Least Absolute Deviation Regression Estimators",

Econometric Theory, 7, 186-199.
Powell (1) Powell, J. (1983), "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569-1575.
Powell (2) Powell, J. (1984), "Least Absolute Deviation Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-25.
Powell (3) Powell, J. (1985), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.
Weiss Weiss, A.A. (1990), "Least Absolute Error Estimation in the Presence of Serial Correlation," *Journal of Econometrics*, 44, 127-158.

Appendix 1: Proofs.

Proof of Lemma ref: lem2 : Equivalently, the statement in Lemma ref: lem2 can be rewritten as follows.

$$\sup_{\|\delta\|<L} \left\| M_T(\hat{\triangle}_2) - M_T(0) + \omega Q_{zz}\hat{\triangle}_2 \right\| = o_p(1)$$

showing that the variations in the set of the $\hat{\triangle}_2$ is embodied in the vector of real numbers δ. By definition, if this statement is true, we have: $\forall \varepsilon > 0, \forall \eta > 0, \exists N > 0, \forall T > N,$

$$P\left[\sup_{\|\delta\|<L} \left\| M_T(\hat{\triangle}_2) - M_T(0) + \omega Q_{zz}\hat{\triangle}_2 \right\| > \eta\right] < \varepsilon$$

The probability function P corresponds to the joint distribution of $(\tilde{\varepsilon}_t, \zeta_t)_{t=1,\ldots,T}$. Let us denote $N_T(\triangle) \equiv M_T(\triangle) - M_T(0) + \omega Q_{zz}\triangle$. The probability can be decomposed as follows.

$$\int 1_{\left[\sup_{\|\delta\|<L\varepsilon} \|N_T(\hat{\triangle}_2)\|>\eta\right]} 1_{\left[\|\hat{\triangle}_2\|<L_\varepsilon\right]} dP + \int 1_{\left[\sup_{\|\delta\|<L\varepsilon} \|N_T(\hat{\triangle}_2)\|>\eta\right]} 1_{\left[\|\hat{\triangle}_2\|\geq L_\varepsilon\right]} dP$$

It can be shown that there exists $L_\varepsilon > 0$ suhc that the second term of this expression is bounded by $P\left[\|\hat{\triangle}_2\| \geq L_\varepsilon\right]$ that can be made smaller than ε/2 for T large enough as soon as $\|\delta\| < L$, because ζ_t is $O_p(1)$ by hypothesis. On the other hand, the first term can be rewritten by conditioning on $\hat{\triangle}_2$, i.e. on ζ_t. This yields

$$\int 1_{\left[\sup_{\|\delta\|<L} \|N_T(\hat{\triangle}_2)\|>\eta\right]} 1_{\left[\|\hat{\triangle}_2\|<L_\varepsilon\right]} dP = \int A(\hat{\triangle}_2)\, dP_{\hat{\triangle}_2} \qquad \#$$

where $P_{\hat{\triangle}_2}$ is the marginal c.d.f. of $\hat{\triangle}_2$ and

$$A(\hat{\triangle}_2) = \int 1_{\left[\sup_{\|\delta\|<L} \|N_T(\hat{\triangle}_2)\|>\eta\right]} 1_{\left[\|\hat{\triangle}_2\|<L_\varepsilon\right]} dP_{\varepsilon|\hat{\triangle}_2}$$

where $P_{\tilde{\varepsilon}|\hat{\triangle}_2}$ is the c.d.f. of $\tilde{\varepsilon}$ conditionally on $\hat{\triangle}_2$. Here, $\hat{\triangle}_2$ can be considered as a nonrandom value and Lemma ref: lem1 can be applied:

$$A(\hat{\triangle}_2) \leq P\left[\sup_{\|\Delta\|<L} \|N_T(\triangle)\| > \eta\right]$$

where $\triangle$ is non random. Therefore, Lemma ref: lem1 implies that $A(\hat{\triangle}_2)$ can be made smaller than ε/2 with T large enough. Then, the first term of expression ref: a1 , which is the expectation of $A(\hat{\triangle}_2)$, is also smaller than ε/2. This proves Lemma ref: lem2 . *QED*.

Proof of Proposition ref: arbase : We define for $\|\delta\| \leq L_1$

$$\hat{\Delta}_1(\delta) \equiv \delta - (1-q)JT^{1/2}(\hat{\pi} - \pi_0) + JT^{1/2}(\hat{\Pi} - \Pi_0)\gamma_0$$

where $\delta \in R^{G+K_1}$ and $J \equiv H(\Pi_0)'[H(\Pi_0)H(\Pi_0)']^{-1}$. Lemma ref: lem2 implies that

$$\sup_{\|\delta\|\leq L_1} \|M_T(\hat{\Delta}_1(\delta)) - M_T(0) + \omega Q_{zz}\hat{\Delta}_1(\delta)\| = o_p(1) \qquad \#$$

for any $L_1 > 0$. Next, we define $\hat{\Delta} \equiv T^{1/2}(\hat{\alpha} - \alpha_0)$. Then, one can show:

$$M_T(\hat{\Delta}_1(\hat{\Delta})) = o_p(1). \qquad \#$$

because $T^{1/2}M_T(\hat{\Delta}_1(\hat{\Delta})) = H(\Pi_0)'[H(\hat{\Pi})H(\hat{\Pi})']^{-1}H(\hat{\Pi})\left[\frac{\partial S_T}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}}\right]_-$. Here, $H(\hat{\Pi})$ is bounded in probability: $H(\hat{\Pi}) = O_p(1)$ by Assumption 2 and $\left[\frac{\partial S_T}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}}\right]_-$ the vector of left hand side partial derivatives of the objective function in ref: a3 , can be shown using Lemma A.1 in Jure kova (1977) to be $o_p(1)$. In a sense, equation ref: a3 expresses that $\hat{\alpha}$ is the extremum estimator associated to our problem. Hence, one can show using Lemma 5.2 in Jure kova (1977) that the results in ref: a2 and ref: a3 together imply that

$$\hat{\Delta} = O_p(1). \qquad \#$$

The final step in deriving the asymptotic distribution of the 2SQR(θ, q) estimator $\hat{\alpha}$ is to combine the results in ref: a2 and ref: a4 to obtain

$$\omega Q_{zz}\hat{\Delta}_1(\hat{\Delta}) = M_T(0) + o_p(1)$$

since $M_T(\hat{\Delta}_1(\hat{\Delta})) = o_p(1)$. By rearranging terms and noting that (i) $Q_{zz}J = H(\Pi_0)'Q$ and (ii) $\psi_\theta(qv_t) = \psi_\theta(v_t)$, we have the asymptotic representation for the 2SQR(θ, q)

$$T^{1/2}(\hat{\alpha} - \alpha_0) = Q_{zz}^{-1}H(\Pi_0)'\{T^{-1/2}\sum_{t=1}^{T} X_t qf(0)^{-1}\psi_\theta(v_t)$$
$$+ (1-q)QT^{1/2}(\hat{\pi} - \pi_0) - QT^{1/2}(\hat{\Pi} - \Pi_0)\gamma_0\} + o_p(1) \qquad QED.$$

Proof of Proposition ref: arbias :

The proof of Proposition ref: arbias is similar to that of Proposition ref: arbase with $\hat{\Delta}_1(\delta) \equiv \delta - (1-q)JT^{1/2}(\hat{\pi} - \pi_0 - B_\pi) + JT^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0$ and $\hat{\Delta} \equiv T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$. QED.

Proof of Proposition ref: normqr : Lemma ref: arqr shows the asymptotic representation for the quantile estimator $T^{1/2}(\hat{\pi}-\pi_0)$ and the asymptotic representation of $T^{1/2}(\hat{\Pi}-\Pi_0)\gamma_0$.

$$T^{1/2}(\hat{\pi}-\pi_0) = Q^{-1}T^{-1/2}\sum_{t=1}^{T}X_t f(0)^{-1}\psi_\theta(v_t)+o_p(1)$$

$$T^{1/2}(\hat{\Pi}-\Pi_0)\gamma_0 = Q^{-1}T^{-1/2}\sum_{t=1}^{T}X_t\xi_t+o_p(1)$$

where ξ_t was defined in Assumption 8. Substituting with these representations into the asymptotic representation for the 2SQR(θ,q) estimator in Proposition ref: arbase and collecting terms give:

$$T^{1/2}(\hat{\alpha}-\alpha_0) = Q_{zz}^{-1}H(\Pi_0)'T^{-1/2}\sum_{t=1}^{T}X_t\eta_t+o_p(1)$$

where $\eta_t \equiv f(0)^{-1}\psi_\theta(v_t)-\xi_t$. By applying the Liapounov's central limit theorem to $T^{-1/2}\sum_{t=1}^{T}X_t\eta_t$ using Assumptions 3,5-6,5$'$-6$'$,7-8, we obtain:

$T^{-1/2}\sum_{t=1}^{T}X_t\eta_t \overset{d}{\to} N(0,\sigma_0^2Q^{-1})$. Therefore, we have $T^{1/2}(\hat{\alpha}-\alpha_0) \overset{d}{\to} N(0,C)$ where $C \equiv \sigma_0^2 Q_{zz}^{-1}$ and $\sigma_0^2 \equiv E(\eta_t)^2$. *QED.*

Proof of Lemma ref: cvsigma :

We consider each term of in the definition of $\hat{\sigma}^2$.

(1) The vector $\hat{\gamma}$ is the first G elements in $\hat{\alpha}$. Hence, $\hat{\gamma}-\gamma_0 = o_p(1)$.

(2) Proof of $\hat{\delta}_{ij}-\delta_{ij} = o_p(1)$.

Consider $\left|T^{-1}\sum_{t=1}^{T}1_{[\hat{V}_{it}\le 0]}1_{[\hat{V}_{jt}\le 0]}-T^{-1}\sum_{t=1}^{T}1_{[V_{it}\le 0]}1_{[V_{jt}\le 0]}\right| \le$
$T^{-1}\sum_{t=1}^{T}\left|1_{[\hat{V}_{it}\le 0]}-1_{[V_{it}\le 0]}\right| + T^{-1}\sum_{t=1}^{T}\left|1_{[\hat{V}_{jt}\le 0]}-1_{[V_{jt}\le 0]}\right|$
$\le T^{-1}\sum_{t=1}^{T}1_{[|V_{it}|<\|X_t'\|\|\hat{\Pi}_i-\Pi_{0i}\|]} + T^{-1}\sum_{t=1}^{T}1_{[|V_{jt}|<\|X_j'\|\|\hat{\Pi}_j-\Pi_{0j}\|]}$
$\equiv Z_{1T}+Z_{2T}$.

The second inequality is obtained using $|1_{[x\le 0]}-1_{[y\le 0]}| \le 1_{[|x|\le|x-y|]}$ and the Cauchy-Schwarz inequality. We only show that $Z_{1T}=o_p(1)$ since the same argument can be applied to Z_{2T}. Let $\epsilon > 0$ and $A \equiv \{Z_{1T} > \epsilon\}$. Consider $\Pr(A) \le \Pr(A\cap B)+\Pr(B^c)$ for any event B. Let $B \equiv \{\|\hat{\Pi}_i-\Pi_{0i}\| \le z\}$ for some $z > 0$. Then, by the consistency of $\hat{\Pi}_i$, $\Pr(B^c) \to 0$. Next we consider $\Pr(A\cap B)$

$\le \Pr(T^{-1}\sum_{t=1}^{T}1_{[|V_{it}|<\|X_t'\|z]} > \epsilon)$
$\le \frac{1}{\epsilon T}\sum_{t=1}^{T}E(1_{[|V_{it}|<\|X_t'\|z]})$ [by the Markov inequality]
$\le \frac{1}{\epsilon T}\sum_{t=1}^{T}\int_{-z\triangle}^{z\triangle}g_{0i}d\lambda$ [by Assumption 10]
$\le \frac{2z\triangle g_{0i}}{\epsilon T}$.

We can choose z arbitrarily small, which implies that $Z_{1T}=o_p(1)$ and similarly $Z_{2T}=o_p(1)$.

Therefore, $\hat{\delta}_{ij}-\delta_{ij} = o_p(1)$.

(3) Proof of $\hat{\delta}_i-\delta_i = o_p(1)$.

Consider $\left|T^{-1}\sum_{t=1}^{T}1_{[\hat{v}_t\le 0]}1_{[\hat{V}_{it}\le 0]}-T^{-1}\sum_{t=1}^{T}1_{[v_t\le 0]}1_{[V_{it}\le 0]}\right| \le$
$T^{-1}\sum_{t=1}^{T}|1_{[\hat{v}_t\le 0]}-1_{[v_t\le 0]}|+T^{-1}\sum_{t=1}^{T}\left|1_{[\hat{V}_{it}\le 0]}-1_{[V_{it}\le 0]}\right| \le$
$T^{-1}\sum_{t=1}^{T}1_{[|v_t|<\|X_t'\|\|\hat{\pi}_i-\pi_0\|]}+T^{-1}\sum_{t=1}^{T}1_{[|V_{it}|<\|X_j'\|\|\hat{\Pi}_i-\Pi_{0i}\|]} \equiv Z_{3T}+Z_{4T}$. By using the same argument as in (2), it can be shown that $Z_{3T}=o_p(1)$ and $Z_{4T}=o_p(1)$. Therefore, $\hat{\delta}_i-\delta_i = o_p(1)$. *QED.*

Appendix 2: Simulations Design

The system is given by

$$B\begin{bmatrix} y_t' \\ Y_t' \end{bmatrix} + \Gamma X_t' = U_t' \qquad \#$$

where $\begin{bmatrix} y_t' \\ Y_t' \end{bmatrix}$ is a 2×1 vector of endogenous variables, X_t' is a 3×1 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 $\Gamma \equiv \begin{bmatrix} -1 & -0.2 & 0 \\ -1 & 0 & -0.4 \end{bmatrix}$. We are interested in the first equation of the system and the system is exactly identified by the zero restrictions $\Gamma_{13} = \Gamma_{22} = 0$. The structural equation in (ref: seqn) can be written in a matrix representation:

$$\begin{bmatrix} y & Y \end{bmatrix} B' = -X\Gamma' + U$$

which gives the following reduced form equations

$$\begin{bmatrix} y & Y \end{bmatrix} = X\begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} + \begin{bmatrix} v & V \end{bmatrix} \qquad \#$$

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 $\pi_0' = (2.3, 0.3, 0.3)$ and $\Pi_0' = (2.6, 0.2, 0.6)$.

The errors $\begin{bmatrix} v & V \end{bmatrix}$ in the reduced form equations in (ref: reqn) are generated in such a way that Assumption 6 and $6'$ are satisfied:

$$v = v^e - F_{v^e}^{-1}(\theta)$$
$$V = V^e - F_{V^e}^{-1}(\theta)$$

where $v^e \sim N(0, I_T)$, $V^e \sim N(0, I_T)$ 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 and third columns in X are also generated using the normal distribution $\mathrm{N}\left(\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$. Since we have assumed that X is not random in the theory, we generate X only once and use the same values for all simulation experiments. 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 (ref: reqn). Here, the parameters in eq. ref: eq0 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 One Step Quantile Regression estimator without correcting for the endogeneity problem is

$$\tilde{\alpha} \in \arg\min \sum_{t=1}^{T} \rho_\theta(y_t - Z_t'\alpha)$$

and $\hat{\alpha}$ is the 2SQR(θ, q) estimator.

The traditional simplex algorithm used to compute the quantile estimator is not efficient because the number of kinks of the objective function grows along with the number of observations. Instead, we use an improved version of simplex algorithm proposed by Barrodale and Roberts (1974). We set the number of replications in all experiments to 1,000 and the number of observation to 50, 100 and 300. We have chosen 5 values (0.05, 0.25, 0.50, 0.75, 0.95) for θ.

Table 1. Means and Standard Deviations of One Step Quantile Estimator

	θ	0.05	0.25	0.50	0.75	0.95
T = 50						
$\tilde{\gamma}$	Mean	0.1131	0.1043	0.1020	0.1070	0.1142
	Std	0.2879	0.1812	0.1708	0.1736	0.2841
$\tilde{\beta}_0$	Mean	1.9226	2.0333	2.1172	2.1808	2.2462
	Std	1.3145	0.6586	0.5264	0.4370	0.5007
$\tilde{\beta}_1$	Mean	0.4249	0.4374	0.4404	0.4353	0.4398
	Std	0.3598	0.2267	0.2110	0.2229	0.3548
T = 100						
$\tilde{\gamma}$	Mean	0.1158	0.1187	0.1264	0.1293	0.1242
	Std	0.1948	0.1241	0.1099	0.1198	0.1913
$\tilde{\beta}_0$	Mean	1.9641	2.0793	2.1514	2.2453	2.3990
	Std	0.9560	0.4793	0.3672	0.3206	0.3726
$\tilde{\beta}_1$	Mean	0.4310	0.4184	0.4107	0.4099	0.3982
	Std	0.2302	0.1549	0.1429	0.1521	0.2299
T = 300						
$\tilde{\gamma}$	Mean	0.1085	0.1135	0.1111	0.1120	0.1103
	Std	0.1087	0.0714	0.0670	0.0726	0.1098
$\tilde{\beta}_0$	Mean	1.9891	2.0940	2.1937	2.2838	2.4147
	Std	0.5351	0.2766	0.2188	0.1934	0.2129
$\tilde{\beta}_1$	Mean	0.3963	0.3903	0.3951	0.3940	0.3931
	Std	0.1236	0.0844	0.0759	0.0858	0.1294

Note: True Structural Parameters: $\gamma_0 = 0.5, \beta_{00} = 1, \beta_{10} = 0.2$

Table 2. Means and Standard Deviations of $2SQR(q,\theta)$ Based on Quantile Prediction

	θ	0.05	0.25	0.50	0.75	0.95
T = 50						
$\hat{\gamma}$	Mean	-3.2926	0.7372	0.5570	0.5699	0.7038
	Std	119.6657	3.4276	1.1030	0.7669	3.5433
$\hat{\beta}_0$	Mean	14.1242	0.3456	0.8310	0.7923	0.3493
	Std	412.5810	9.6412	3.1042	2.2001	10.4769
$\hat{\beta}_1$	Mean	1.8155	0.0487	0.1792	0.1598	0.0799
	Std	49.6428	2.0736	0.7927	0.5350	2.1024
T = 100						
$\hat{\gamma}$	Mean	0.5741	0.5285	0.5199	0.5336	0.5332
	Std	0.9476	0.2983	0.266	0.305	1.5451
$\hat{\beta}_0$	Mean	0.7929	0.9205	0.9381	0.8878	0.8941
	Std	2.8897	0.9209	0.8346	0.9504	4.9271
$\hat{\beta}_1$	Mean	0.1465	0.1801	0.1852	0.1807	0.1781
	Std	0.6296	0.2288	0.2045	0.2468	0.8870
T = 300						
$\hat{\gamma}$	Mean	0.5331	0.5119	0.5003	0.5056	0.5049
	Std	0.3428	0.1667	0.1505	0.1653	0.2888
$\hat{\beta}_0$	Mean	0.9012	0.9598	0.9978	0.9786	0.9775
	Std	1.0495	0.5264	0.4725	0.5197	0.9161
$\hat{\beta}_1$	Mean	0.1823	0.1976	0.2000	0.1994	0.2027
	Std	0.2220	0.1172	0.1059	0.1150	0.1915

Table 3. Means and Standard Deviations of $2SQR(q,\mathrm{q})$ Based on LS Prediction (q = 0.1)

	q	0.05	0.25	0.50	0.75	0.95
T = 50						
$\hat{g}$	Mean	0.5409	0.5392	0.5399	0.5390	0.5396
	Std	0.3804	0.3776	0.3747	0.3712	0.3829
$\hat{b}_0$	Mean	1.4795	1.1306	0.8835	0.6413	0.2845
	Std	1.7153	1.3407	1.0810	0.8264	0.4988
$\hat{b}_1$	Mean	0.1711	0.1729	0.1732	0.1741	0.1751
	Std	0.2848	0.2811	0.2804	0.2786	0.2824
T = 100						
$\hat{g}$	Mean	0.5130	0.5121	0.5127	0.5125	0.5107
	Std	0.2065	0.2042	0.2048	0.2056	0.2062
$\hat{b}_0$	Mean	1.5989	1.2254	0.9607	0.6992	0.3237
	Std	0.9801	0.7730	0.6406	0.5070	0.3203
$\hat{b}_1$	Mean	0.1899	0.1896	0.1893	0.1901	0.1908
	Std	0.1631	0.1613	0.1613	0.1621	0.1610
T = 300						
$\hat{g}$	Mean	0.4990	0.4999	0.4992	0.4994	0.4972
	Std	0.1210	0.1182	0.1182	0.1181	0.1190
$\hat{b}_0$	Mean	1.6609	1.2675	1.0000	0.7288	0.3431
	Std	0.5793	0.4524	0.3738	0.2956	0.1876
$\hat{b}_1$	Mean	0.2018	0.2015	0.2017	0.2016	0.2027
	Std	0.0840	0.0827	0.0826	0.0826	0.0827

Table 4. Means and Standard Deviations of $2SQR(q,\mathrm{q})$ Based on LS Prediction (q = 0.5)

	q	0.05	0.25	0.50	0.75	0.95
T = 50						
$\hat{g}$	Mean	0.5457	0.5375	0.5410	0.5362	0.5392
	Std	0.4759	0.4105	0.3913	0.3825	0.4923
$\hat{b}_0$	Mean	0.8314	0.8737	0.8803	0.9111	0.9139
	Std	2.1444	1.4594	1.1287	0.8524	0.6433
$\hat{b}_1$	Mean	0.1632	0.1721	0.1733	0.1780	0.1827
	Std	0.3609	0.3032	0.2951	0.2912	0.3534
T = 100						
$\hat{g}$	Mean	0.5156	0.5111	0.5143	0.5130	0.5040
	Std	0.2663	0.2171	0.2159	0.2229	0.2630
$\hat{b}_0$	Mean	0.9388	0.9641	0.9563	0.9644	0.9800
	Std	1.2632	0.8196	0.6771	0.5488	0.4149
$\hat{b}_1$	Mean	0.1907	0.1893	0.1878	0.1916	0.1951
	Std	0.2106	0.1739	0.1703	0.1772	0.2023
T = 300						
$\hat{g}$	Mean	0.4979	0.5026	0.4990	0.4999	0.4891
	Std	0.1647	0.1277	0.1251	0.1267	0.1551
$\hat{b}_0$	Mean	1.0106	0.9872	1.0007	0.9961	1.0110
	Std	0.7867	0.4881	0.3952	0.3172	0.2436
$\hat{b}_1$	Mean	0.2023	0.2009	0.2017	0.2016	0.2067
	Std	0.1133	0.0906	0.0877	0.0894	0.1086

Table 5. Means and Standard Deviations of $2SQR(q,\mathrm{q})$ Based on LS Prediction (q = 1)

	q	0.05	0.25	0.50	0.75	0.95
T = 50						
$\hat{g}$	Mean	0.5518	0.5354	0.5424	0.5327	0.5388
	Std	0.6860	0.4865	0.4406	0.4428	0.7203
$\hat{b}_0$	Mean	0.0213	0.5525	0.8763	1.2483	1.7006
	Std	3.0918	1.7303	1.2702	0.9891	0.9511
$\hat{b}_1$	Mean	0.1533	0.1710	0.1736	0.1828	0.1923
	Std	0.5251	0.3602	0.3357	0.3390	0.5187
T = 100						
$\hat{g}$	Mean	0.5189	0.5098	0.5164	0.5136	0.4957
	Std	0.4021	0.2586	0.2473	0.2664	0.3943
$\hat{b}_0$	Mean	0.1136	0.6374	0.9509	1.2959	1.8003
	Std	1.9094	0.9729	0.7775	0.6558	0.6264
$\hat{b}_1$	Mean	0.1917	0.1889	0.1858	0.1934	0.2006
	Std	0.3156	0.2091	0.1945	0.2136	0.3046
T = 300						
$\hat{g}$	Mean	0.4966	0.5059	0.4989	0.5007	0.4790
	Std	0.2531	0.1540	0.1441	0.1517	0.2373
$\hat{b}_0$	Mean	0.1976	0.6369	1.0017	1.3301	1.8459
	Std	1.2074	0.5875	0.4547	0.3793	0.3714
$\hat{b}_1$	Mean	0.2029	0.2002	0.2016	0.2015	0.2118
	Std	0.1741	0.1107	0.1012	0.1075	0.1685