amse
aix marseille
school of economics

Working Papers / Documents de travail

Bias Transmission and Variance Reduction in Two-Stage Quantile Regression

Tae-Hwan Kim, Yonsei University, School of Economics
Christophe Muller, Aix-Marseille Université, Greqam

WP 2012 - Nr 21

Aix*Marseille université · cnrs · L'ECOLE DES HAUTES ETUDES EN SCIENCES SOCIALES · Inserm · IRD Institut de recherche pour le développement · Centrale Marseille

Bias Transmission and Variance Reduction in Two-Stage Quantile Regression

Tae-Hwan Kim[a,*] and Christophe Muller[b]

[a]School of Economics, Yonsei University, Seoul 120-749, Korea
tae-hwan.kim@yonsei.ac.kr

[b]University of Aix-Marseille (Aix-Marseille School of Economics), 14, avenue Jules Ferry, F-13621 Aix-en-Provence cedex, France.
christophe.muller@univ-amu.fr

July 2012

Abstract:

In this paper, we propose a variance reduction method for quantile regressions with endogeneity problems. First, we derive the asymptotic distribution of two-stage quantile estimators based on the fitted-value approach under very general conditions on both error terms and exogenous variables. Second, we exhibit a bias transmission property derived from the asymptotic representation of our estimator. Third, using a reformulation of the dependent variable, we improve the efficiency of the two-stage quantile estimators by exploiting a trade-off between an asymptotic bias confined to the intercept estimator and a reduction of the variance of the slope estimator. Monte Carlo simulation results show the excellent performance of our approach. In particular, by combining quantile regressions with first-stage trimmed least-squares estimators, we obtain more accurate slope estimates than 2SLS, 2SLAD and other estimators for a broad range of distributions.

Key words: Two-Stage Estimation, Variance Reduction, Quantile Regression, Asymptotic Bias.
JEL codes: C13, C30.

*We are grateful to R. Davidson, W. Newey, J. Powell, H. White, O. Linton, C.B. Phillips, S. Portnoy, R. Blundell and P. Lavergne for useful discussions and to participants in presentations in the University of Nottingham, CREST-INSEE in Paris, London School of Economics, University of California at San Diego, University of Alicante, University of Aix-Marseille, University of Cergy-Pontoise, Toulouse School of Economics, Yonsei University and several conferences for their comments. Remaining errors are ours. We acknowledge the Award No. 30,649 from the British Academy. Tae-Hwan Kim is grateful for the financial support from the National Research Foundation of Korea – a grant funded by the Korean Government (NRF-2009-327-B00088). Christophe Muller is grateful for the financial support by Spanish Ministry of Sciences and Technology, Project No. SEJ2005-02829/ECON.

1 Introduction

This paper considers the estimation of a structural equation using quantile regression. Since the seminal work by Koenker and Bassett (1978), the literature on quantile regression has grown rapidly. There are two strands in the literature about quantile regression in the presence of endogeneity. The first one, which we call the 'structural approach,' corresponds to models specified in terms of the conditional quantile function of the structural equation and usually allows for heterogeneous (or non-constant) quantile treatment effects through nonseparable models.[1] On the other hand, the second one, which we refer to as the 'fitted-value approach,' is based on the conditional quantile function of the reduced-from equation. In the latter approach, the analysts substitute the endogenous regressors with their fitted values obtained from some auxiliary regression based on other exogenous variables.We follow the latter approach in this paper. For quantile regressions, it is anchored on conditional quantile restrictions applied to the reduced-form equation.

Owing to its computational convenience, the fitted-value approach has been used to estimate homogeneous (or constant) quantile treatment effects.[2] The theoretical foundation for the fitted-value approach was first laid by Amemiya (1982) and Powell (1983) who analyze the two-stage least-absolute-deviations estimator in a simple setting. Chen (1988) and Chen and Portnoy (1996) investigate two-stage quantile regression in which the trimmed least squares (TLS) and least absolute deviations (LAD) estimators are employed as the first-stage estimators, assuming that the error terms are independent and identically distributed (IID). Kim and Muller (2004) use a similar approach with quantile regression in the first stage. Although the fitted-value approach has sometimes been used in applied work in which data are affected by serial correlation and heteroskedasticity, no theoretical results under such general conditions have been provided.

In this paper, we make several contributions to the literature on the fitted-value approach. First, we derive the asymptotic distribution and the variance-covariance matrix of the two-stage quantile estimator under very general conditions on both error terms and explanatory variables. Second, we exhibit a 'bias transmission property' that characterizes the asymptotic representation of our estimator. We use this property to facilitate the analyses of the link of reduced-form and structural model, and to confine estimation bias on the intercept for some models. Third, we propose a new method to improve the efficiency of the two-stage quantile regressions. This method is based on an idea in Amemiya (1982), who uses a composite dependent variable that is a weighted combination of the original dependent variable and its fitted value. This composite dependent variable is used in the second stage and the combination weight is varied to reduce the variance of the two-stage quantile estimator.

A well-known method to reduce the variance of an estimator in statistics is to combine the estimator with either a fixed point or another estimator, which creates the trade-off between bias and efficiency. In this case, the optimal combination weight is determined by minimizing the asymptotic variance of the combination estimator.[3] This approach is difficult to apply to two-stage estimation because it generally either makes the entire estimator biased or requires the estimation of the joint distribution of the two combined estimators. In contrast, our approach of using a composite dependent variable implies that (i) only the intercept estimator is inconsistent with the consistency of the slope estimator not depending on the weight parameter, and (ii) the variance

[1] The literature on the structural approach for quantile regressions is abundant. See for example: Kemp (1999), MaCurdy and Timmins (2000), Sakata (2001), Abadie et al. (2002), Chen et al. (2003), Chesher (2003), Hong and Tamer (2003), Honore and Hu (2003), Chernozhukov and Hansen (2005, 2006, 2008), Imbens and Newey (2006), Ma and Koenker (2006), Chernozhukov, Imbens and Newey (2007), Lee (2007).

[2] For example, Arias et al. (2001), Garcia et al. (2001), Chevapatrakul et al. (2009) and Chortareas et al. (2012).

[3] For example, see James and Stein (1960), Sen and Saleh (1987), and Kim and White (2001).

of the consistent slope estimator can be reduced by adjusting the weight parameter. The results of our Monte Carlo simulations show that considerable efficiency gains can be achieved by using the proposed variance reduction method. This is particularly noticeable when compared to typical estimators such as 2SLS or 2SLAD.

The paper is organized as follows. Section 2 discusses the model and the assumptions. In Section 3, we derive the asymptotic representation of the two-stage quantile regression estimator. We characterize the asymptotic bias of general two-stage estimators in Section 4. We analyze in Section 5 the asymptotic normality and the asymptotic covariance matrix of two-stage quantile estimators based on LS or TLS predictions. We also investigate the optimal weights that minimize the asymptotic variance of two-stage quantile estimators. In Section 6, we present Monte Carlo simulation results. Finally, Section 7 concludes. All technical proofs are collected in Appendix A.

2 The Model

We are interested in estimating the parameter (α_0) in the following structural equation by quantile regression:

$$\begin{aligned} y_t &= x_{1t}'\beta_0 + Y_t'\gamma_0 + u_t \\ &= z_t'\alpha_0 + u_t, \end{aligned} \tag{1}$$

where $[y_t, Y_t']$ is a $(G+1)$ row vector of endogenous variables, x_{1t}' is a K_1 row vector of exogenous variables, $z_t = [x_{1t}', Y_t']'$, $\alpha_0 = [\beta_0', \gamma_0']'$ and u_t is an error term. We denote by x_{2t}' the row vector of the K_2 exogenous variables excluded from (1).

By assumption, the first element of x_{1t} is 1 [4]. This crucial assumption will allow us to confine a bias to an intercept parameter, often less interesting for analysts than the slope coefficients. We further assume that Y_t can be linearly predicted from all of the exogenous variables:

$$Y_t' = x_t'\Pi_0 + V_t', \tag{2}$$

where $x_t' = [x_{1t}', x_{2t}']$ is a K row vector with $K = K_1 + K_2$, Π_0 is a $K \times G$ matrix of unknown parameters and V_t' is a G row vector of unknown error terms.

Using (1) and (2), y_t can also be expressed as follows:

$$y_t = x_t'\pi_0 + v_t, \tag{3}$$

where

$$\begin{aligned} \pi_0 &= H(\Pi_0)\alpha_0 \text{ with } H(\Pi_0) = \left[\begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi_0 \right] \\ \text{and } v_t &= u_t + V_t'\gamma_0. \end{aligned} \tag{4}$$

Equations (2) and (3) are the basis of the first-stage estimation that yields estimators $\hat{\pi}$ and $\hat{\Pi}$ respectively of π_0 and Π_0. So far, we did not mention any restriction on errors. The precise error restrictions will be introduced below in Assumptions 3 and 4 when dealing with examples of first-stage estimators. This is because we wish to keep the framework as general as possible until we deal with these examples. However, to set ideas, the reader may wish to consider conditional quantile restriction on v_t in the fitted-value approach. We now specify the data generating process.

[4] Using another coefficient of secondary interest is possible for our argument, while it makes the presentation more interesting by choosing the intercept.

Assumption 1. The sequence $\{(x_t', u_t, v_t)\}$ is $\alpha-$mixing with mixing numbers $\{\alpha(s)\}$ of size $-2\,(4K+1)\,(K+1)$.[5]

Studying quantile regressions with α-mixing processes is an unusual degree of generality. One step in this direction was made by Portnoy (1991), who derived asymptotic results of quantile estimators in dependent and even non-stationary cases, using $m(n)$-decomposability of random variables.

It is generally possible to employ unbiased estimators in the first stage. However, in order to exploit later on a trade-off between bias and efficiency, we allow in Assumption 2 for inconsistent first-stage estimation with bounded asymptotic bias terms in the following assumption. This form is convenient for including the contribution of first-stage estimators in the asymptotic distribution of our final estimator. The precise restrictions on v_t and V_t corresponding to π_0 and Π_0 will be brought up later on.

Assumption 2. There exist finite bias vectors B_π and B_Π such that
$T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) = O_p(1)$ and $T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi) = O_p(1)$.

When the bias terms B_π and B_Π are zero, the first-stage estimators are consistent. A case of non-zero bias terms is when the reduced form equation in (3) is estimated by LS to produce the first-stage estimator $\hat{\pi}$, whereas the usual conditional quantile restriction (i.e. the zero quantile restriction) is placed on v_t in the same equation. In that case, the zero mean restriction on v_t cannot be simultaneously satisfied in general. This implies that the intercept estimator, at least in $\hat{\pi}$, is inconsistent.

Let us now say more about two-stage quantile regressions in our setting. For any quantile $\theta \in (0,1)$, we define $\rho_\theta(z) = z\psi_\theta(z)$, where $\psi_\theta(z) = \theta - 1_{[z\leq 0]}$ and $1_{[.]}$ is the indicator function. If the orthogonality conditions, $E(z_t\psi_\theta(u_t)) = 0$, were satisfied, then the usual one-stage quantile regression estimator (QR) would be consistent. However, when u_t and Y_t (a sub-vector of z_t) are statistically linked under weak endogeneity of Y_t, these conditions may not be satisfied. In that case, the QR of α_0 is generally not consistent, which is the endogeneity problem that prevents us from using simple quantile regressions.

As an extension of Amemiya (1982), Powell (1983) and Chen and Portnoy (1996) to broader DGPs, we define, for any quantile θ, the Two-Stage Quantile Regression (2SQR(θ, q)) estimator $\hat{\alpha}$ of α_0 as a solution to the following program:

$$\min_{\alpha} \;\; S_T(\alpha, \hat{\pi}, \hat{\Pi}, q, \theta) = \sum_{t=1}^{T} \rho_\theta(qy_t + (1-q)\hat{y}_t - x_t' H(\hat{\Pi})\alpha), \tag{5}$$

where $\hat{y}_t = x_t'\hat{\pi}$, q is a positive scalar constant, and $\hat{\pi}, \hat{\Pi}$ are first-stage estimators. In the quantile regression in (5), the dependent variable $qy_t + (1-q)\hat{y}_t$ is a weighted average of y_t and of its fitted-value $\hat{y}_t$ obtained from the reduced form equation in (3). The combination weight q is restricted to be positive for a technical reason discussed in the proof of Proposition 1 below. Alternatively, as in Powell (1983), the case q negative is also possible by imposing $\theta = 0.5$, i.e., with the LAD estimator.

The reformulation of the dependent variable as $qy_t + (1-q)\hat{y}_t$ was originally suggested by Amemiya (1982) to improve efficiency in two-stage estimation with $0 < q < 1$. The case $q =$

[5] The sequence $\{W_t\}$ of random variables is $\alpha-$mixing if $\alpha(s)$ decreases towards 0 as $s \to \infty$, where $\alpha(s) = \sup_t \sup_{A\in F_{-\infty}^t; B\in F_{t+s}^{\infty}} |P(A\cap B) - P(A)P(B)|$ for $s \geq 1$ and F_s^t denote the σ-field generated by $(W_s, \ldots, W_t)$ for $-\infty \leq s \leq t \leq \infty$. The sequence is called $\alpha-$mixing of size $-a$ if $\alpha(s) = O(s^{-a-\varepsilon})$ for some $\varepsilon > 0$.

1 corresponds to the usual two-stage quantile regression estimator, while $q = 0$ corresponds to the inverse regression estimator under exact identification. Thus, the new dependent variable introduces a trade-off between two estimation methods. To obtain the asymptotic distribution of the 2SQR(θ, q) estimator, we impose the following usual regularity conditions.

Assumption 3. (i) $H(\Pi_0 + B_\Pi)$ is of full column rank.
(ii) Let $F_t(.|x)$ be the conditional cumulative distribution function (CDF) and $f_t(.|x)$ be the conditional probability density function (PDF) of v_t. The conditional PDF $f_t(\cdot|x)$ is assumed to be Lipschitz continuous for all x, strictly positive and bounded by a constant f_0 (i.e., $f_t(\cdot|x) < f_0$, for all x).
(iii) The matrices $Q = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]$ and $Q_0 = \lim_{T\to\infty} E\left[\frac{1}{T}\sum_{t=1}^{T} f_t(0|x_t) x_t x_t'\right]$ are finite and positive definite.
(iv) $E(\psi_\theta(v_t)|x_t) = 0$, for an arbitrary θ.
(v) There exists a positive number $C > 0$ such that $E(\|x_t\|^3) < C < \infty$ for any t.

Assumption 3(i) is analogous to the usual identification condition for simultaneous equations models. The bias B_Π appears in the condition because the first-stage estimator converges towards $\Pi_0 + B_\Pi$. In the case when OLS is used for estimating Π_0, Assumption 3(i) ensures that $E[x_t Y_t] \neq 0$. It implies similar conditions when other estimators are used. Assumption 3(ii) simplifies the demonstration of convergence of remainder terms to zero for the calculation of the asymptotic representation. The second part of Assumption 3(iii) is the counterpart of the usual condition for OLS that the sample second moment matrix of the regressor vectors converges towards a finite positive definite matrix, which corresponds to the first part. The last condition is akin to the one in the conventional IV approach in that this condition is necessary for consistency and for the inversion of the relevant empirical process to establish the asymptotic normality.

Assumption 3(iv) is the assumption that zero is the given θ^{th}-quantile of the conditional distribution of v_t.[6] It identifies the coefficients of the model. Assumption 3(v), the moment condition on the exogenous variables, is necessary for the stochastic equicontinuity of our empirical process in the dependent case, which is used for the asymptotic representation. We also use it to bound the asymptotic covariance matrix of the parameter estimators. The conditions on the exogenous regressors are weaker than what is usually employed in the two-stage quantile regression literature.

Assumption 3(iv) is central to our fitted-value approach. The conditional quantile restriction is placed on the reduced-form error v_t and the information set used for the conditional restriction exclusively consists of exogenous variables x_t. It has been used in simpler settings in Amemiya (1982), Powell (1981), Chen and Portnoy (1998) and Kim and Muller (2004).

Given that v_t is a function of the structural error u_t and the prediction equation error V_t, it may be difficult to interpret the quantile restriction in Assumption 3(iv) and to provide good examples in which such a restriction holds. One condition to make Assumption 3(iv) plausible and easy to interpret is the quantile-independence condition: $E(\psi_\theta(v_t)|x_t) = E(\psi_\theta(v_t))$. Under the quantile-independence condition, only constant quantile treatment effect models can be specified and estimated. In that case, the intercept estimates describe the parallel shifts in the unconditional quantiles of the error term.

Despite being limited to constant quantile treatment effect models, the fitted-value approach

[6] In the iid case, the term $f(F^{-1}(\theta))^{-1}$ typically appears in the variance formula of a quantile estimator (Koenker and Bassett, 1978). However, due to Assumption 3(iv), $F^{-1}(\theta)$ is now zero so that we have $f(0)^{-1}$ instead, in this case.

has several advantages worth mentioning. First, it does not involve any computational problem, even with a large number of endogenous and exogenous variables, whereas the structural approach cannot handle such a case. For example, the grid search method in Chernozhukov and Hansen (2006) becomes increasingly difficult to implement as the number of endogenous variables increases. Second, considering the fitted-value approach allows us to propose a new powerful method of variance reduction. We now study the asymptotic properties of 2SQR(θ, q) in the next section.

3 The Asymptotic Representation

To derive the asymptotic representation of the 2SQR(θ, q) estimator, we define the following empirical process.

$$M_T(\Delta) = T^{-1/2} \sum_{t=1}^{T} x_t \psi_\theta(q v_t - T^{-1/2} x_t' \Delta),$$

where Δ is a $K \times 1$ vector. Applying Theorem II.8 in Andrews (1990) yields the following lemma. The lemma is proven only for the quantile regression case, while similar derivations can be done for other two-stage M-estimators.

Lemma 1. Suppose that Assumptions 1 and 3 hold. Then, for any $L > 0$, we have the following result:

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

Combining Lemma 1 and Assumption 2 allows us to obtain the asymptotic representation for the 2SQR(θ, q) estimator with a possible bias term B_α as follows[7].

Proposition 1. Suppose that Assumptions 1-3 hold. Then, the asymptotic representation for the 2SQR(θ, q) estimator is:

$$\begin{aligned} T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) &= R T^{-1/2} \sum_{t=1}^{T} x_t q \psi_\theta(v_t) \\ &\quad + (1-q) R Q_0 T^{1/2} (\hat{\pi} - \pi_0 - B_\pi) \\ &\quad - R Q_0 T^{1/2} (\hat{\Pi} - \Pi_0 - B_\Pi) \gamma_0 + o_p(1), \end{aligned}$$

where $B_\alpha = R Q_0 \{(1-q) B_\pi - B_\Pi \gamma_0\}$, $R = Q_{zz}^{*-1} H(\Pi_0^*)'$, $Q_{zz}^* = H(\Pi_0^*)' Q_0 H(\Pi_0^*)$ and $\Pi_0^* = \Pi_0 + B_\Pi$.

The asymptotic representation of 2SQR(θ, q) is composed of four additive right-hand-side terms. The first term does not perturb consistency under Assumption 3(iv) and corresponds to the contribution of the second stage to the uncertainty of the estimator. The second and third terms correspond to the respective contributions of $\hat{\pi}$ and $\hat{\Pi}$ to this uncertainty. Then, when $\hat{\pi}$ and $\hat{\Pi}$ are consistent, it is straightforward to show that the 2SQR(θ, q) is consistent. If $q = 1$, the influence of $\hat{\pi}$ vanishes. The presence of the contribution of $\hat{\pi}$ may imply contradictions between some chosen restrictions on errors in the first and second stages and cause biases, which will be explained in detail in the next section. The formula of B_α is obtained as the value allowing $T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1)$,

[7] Other derivations of asymptotic representation of quantile regression estimators have been developed (Phillips, 1991, Pollard, 1991), which involve slightly different assumptions. Other possible approach to the asymptotic representation of 2SQR is in Chen et al. (2003).

and is derived from the first-order conditions of the second-stage estimation, which is discussed in detail in Appendix A. We note that the consistency of $\hat{\alpha}$ to $\alpha_0 + B_\alpha$ naturally follows from Proposition 1.

4 The Asymptotic Bias

In this section, we discuss the relationship linking the bias terms in the first stage estimators (B_π, B_Π) and the second stage estimators (B_α). We first state the following proposition.

Proposition 2. Suppose that both B_π and B_Π have non-zero components only for their first K_1 elements; i.e., $B_\pi = \begin{bmatrix} B_{\pi,1} \\ B_{\pi,2} \end{bmatrix}$ and $B_\Pi = \begin{bmatrix} B_{\Pi,1} \\ B_{\Pi,2} \end{bmatrix}$ with $B_{\pi,1}$ and $B_{\Pi,1}$ being the first K_1 components, and $B_{\pi,2} = 0$ and $B_{\Pi,2} = 0$. Under this restriction, we have:

$$B_\alpha = \begin{bmatrix} (1-q)B_{\pi,1} - B_{\Pi,1}\gamma_0 \\ 0_G \end{bmatrix},$$

where 0_G is a G vector of zeros.

The sufficient stochastic assumptions for this characterization of the bias transmission are very general, including general serial correlations and heteroskedasticity.

When the first-stage estimation is performed on the two reduced form equations in (2) and (3), the explanatory or instrumental variables x_t consists of vectors x_{1t} and x_{2t}. If a non-zero asymptotic bias is present only in the coefficients of x_{1t} in the first-stage estimators ($\hat{\pi}$ and $\hat{\Pi}$), which is the case on which we focus, then the non-zero asymptotic bias in the second-stage estimator ($\hat{\alpha} = [\hat{\beta}', \hat{\gamma}']'$) is exclusively confined to the coefficients of x_{1t}; that is, only $\hat{\beta}$ is asymptotically biased. Therefore, the parameter (γ_0) for the endogenous variable Y_t in the structural equation in (1), in which applied economists are usually interested, can be consistently estimated.

This is useful because empirical researchers may often pay little attention to the intercept term, whereas the estimates of the slope coefficients often carry more explanatory meaning. One example of such situation is when one performs the first-stage estimation of (2) and (3) by LS. Indeed, under Assumption 3(iv) or under the stronger quantile-independence condition corresponding to constant quantile treatment effect models, the zero mean condition on v_t is not satisfied. Then, the intercept LS estimator cannot be consistent (unless the probability support of v_t does not include the interval between the mean and the quantile of interest), while the remaining slope coefficients can still be consistently estimated. Moreover, depending on distributional restrictions on V_t that the researcher is willing to impose, she may choose an estimator for Π_0 such that its intercept term only is inconsistent.

We emphasize that the researcher, if she or he wishes, can eliminate the bias completely by choosing $q = 1$ (i.e., not using the composite dependent variable) and by placing some suitable restriction on V_t to make $\hat{\Pi}$ consistent. In this paper, we propose instead choosing $q \neq 1$ and selecting first-stage estimators with bias terms confined only to their intercept parts in order to allow for efficiency gains for the second stage slope estimators. In that case, the slope coefficients in the structural equation can be consistently estimated while its asymptotic variance will depend on q. This generates a trade-off between the bias in the intercept estimator (i.e., the first element of $\hat{\alpha}$) and the efficiency of the slope estimator (i.e., all of the remaining elements of $\hat{\alpha}$). Our approach contrasts with the statistics literature on shrinking estimators where all of coefficients estimators are biased to improve their efficiency.

Let $v_t^* = v_t - F^{-1}_{v_t|x_t}(\theta)$. Consider $E(\psi_\theta(v_t^*)|x_t) = \theta - P\left[v_t^* \leq 0|x_t\right] = \theta - P\left[v_t \leq F^{-1}_{v_t|x_t}(\theta)|x_t\right] = \theta - \theta = 0$, which is the conditional quantile restriction characterising v_t^*. As a consequence, we obtain the reduced-form quantile regression restriction, provided we accept the introduction in the regression of a possible nuisance bias term $F^{-1}_{v_t|x_t}(\theta)$ that may affect all coefficients of the model when it is linear in x_t, or even be nonlinear in x_t. Let us now assume that u_t and V_t are independent of $\tilde{x}_t$, defined as x_t except constant variables. Since $v_t = u_t + V_t\gamma_0$, this assumption implies $F^{-1}_{v_t|x_t}(\theta) = F^{-1}_{v_t}(\theta)$ and the nuisance term is confined to the intercept. Then, according to Proposition 2, a bias is generated exclusively on the intercept term of the structural model.

However, this condition of independence for all θ also implies constant effects in the quantile regressions of interest. Although the above characterisation of instrumental variables may be deemed to be strong by some authors, it is usually the way instrumental variables are intuitively found by empiricists: variables that are not connected at all with the model errors seen as a remainder of the explanation of the dependent variable given the effects of explanatory variables.

It is also easy to see that starting instead from $E(\psi_\theta(v_t)|x_t) = 0$ and assuming the independence of v_t and V_t with respect to $\tilde{x}_t$, we can obtain the structural restriction $E(\psi_\theta(u_t^*)|x_t) = 0$ for a structural model with the right value of parameters, except perhaps for a bias on the intercept that is incorporated in the definition of u_t^*. In that sense, under the previous independence assumption, reduced form quantile regressions and structural quantile regressions can be seen as emerging somehow from equivalent restrictions, with perhaps the exception of the intercept. In particular, under the independence/constant effect hypothesis, the estimates based on the conditional quantile of the reduced form can be used to recover the slope estimator of the conditional quantile of the structural form.

5 Asymptotic Normality and Covariances with LS and Trimmed-LS Predictions

In this section, we examine the use of (non-robust) LS estimation and (robust) trimmed-least-squares (TLS) estimation of π_0 and Π_0 in the first stage in this section. There are several reasons to consider LS estimation in the first stage. First, LS estimation is popular and can be readily used. Second, using LS estimation in the first stage may improve the efficiency of the final two-stage quantile estimator, which may explain why empirical researchers have been using this technique.[8] Alternatively, using TLS in the first stage guarantees the robustness of this estimation stage, while some efficiency may be lost. Using twice the same quantile regression in both stages has been examined in Kim and Muller (2004). In that case, there is no consistency issue, but also no opportunity for asymptotic variance reduction either as the asymptotics is invariant to the choice of parameter q.

Since we consider constant quantile treatment effect models, it is not restrictive to assume that x_t is mean-independent of v_t and V_t as in Assumption 3′ below. The following assumption is more restrictive than what is necessary for Assumption 2 or for Proposition 2, given that we now wish to confine the possible bias to the intercept exclusively.

Assumption 3′. (i) $E(v_t|x_t) = E(v_t)$ and (ii) $E(V_t|x_t) = E(V_t)$.

Assumption 3′ imposes the orthogonality of the reduced form errors with all non-constant

[8] Examples include Arias et al. (2001), Garcia et al. (2001), Chevapatrakul et al. (2009) and Chortareas et al. (2012).

exogenous variables. As stated before, an issue of using LS estimation in the first-stage is that the condition $E(v_t) = 0$, which makes the LS estimator $\hat{\pi}$ in (3) consistent, conflicts with the restriction $E(\psi_\theta(v_t)) = 0$, which is implied by Assumption 3(iv); that is, the θ^{th} quantile and the mean of v_t cannot be zero at the same time. Then, to be able to use the usual Bahadur representation of the LS estimator, we define the centered errors $v_t^* = v_t - E(v_t)$ and $V_t^* = V_t - E(V_t)$. By construction, $E(v_t^*|x_t) = 0$ and $E(V_t^*|x_t) = 0$.

Under Assumption 3′, the reduced form equations for Y_t and y_t in (2) and (3) can be rewritten by reallocating the bias to the intercept coefficient as follows:

$$Y_t' = x_t'\Pi_0^* + V_t'^*, \tag{6}$$

where $\Pi_0^* = \Pi_0 + B_\Pi$ with $B_\Pi = [E(V_t)', 0', ..., 0']'$, which is a $(K \times G)$ matrix, and

$$y_t = x_t'\pi_0^* + v_t^*, \tag{7}$$

where $\pi_0^* = \pi_0 + B_\pi$ with $B_\pi = [E(v_t), 0, ..., 0]'$, which is a $(K \times 1)$ matrix.

The bias B_π is generally non-zero for $q \neq 1$. In contrast, B_Π can be non-zero or not, even with $q = 1$, depending on the restrictions imposed on V_t. In the case $q = 1$, a natural specification suggests $E(V_t|x_t) = 0$ when using OLS to estimate (2) and no bias at all. In other cases, B_π and B_Π may have to be taken into account.

Let $\tilde{\Pi}$ and $\tilde{\pi}$ be the first-stage LS estimators based on (6) and (7) respectively and let $\tilde{\alpha}$ be the corresponding 2SQR(θ, q) estimator. The asymptotic representations of $\tilde{\Pi}$ and $\tilde{\pi}$ are obtained and plugged into the formula in Proposition 1 to obtain the asymptotic representation for $\tilde{\alpha}$ as follows:

$$\begin{aligned} T^{1/2}(\tilde{\alpha} - \alpha_0 - B_\alpha) &= RT^{-1/2}\sum_{t=1}^{T} x_t q\psi_\theta(v_t) \\ &\quad -RQ_0Q^{-1}T^{-1/2}\sum_{t=1}^{T} x_t(qv_t^* - u_t^*) + o_p(1). \end{aligned}$$

Due to the characterization of B_α in Proposition 2, we have $B_\alpha = ((1-q)E(v_t) - E(V_t')\gamma_0, 0, \ldots, 0)'$. The intercept estimator is inconsistent, while the slope estimators are not. To derive the asymptotic normality of $\tilde{\alpha}$, we impose the following additional regularity assumptions.

Assumption 4. (i) There exist finite constants Δ_v and Δ_{V_j} such that $E|x_{ti}v_t^*|^3 < \Delta_v$ and $E|x_{ti}V_{jt}^*|^3 < \Delta_{V_j}$, for all i, j and t.
(ii) The covariance matrix $V_T = var\left(T^{-1/2}\sum_{t=1}^{T} S_t\right)$ is positive definite for T sufficiently large, where $S_t = (q\psi_\theta(v_t), qv_t^* - u_t^*)' \otimes x_t, u_t^* = v_t^* - V_t^{*'}\gamma_0$ and $\otimes$ is the Kronecker product.

Assumption 4(i) is used to apply a CLT appropriate for the α-mixing case. It can be much relaxed in the iid case. Assumption 4(ii) ensures the positive definiteness of the variance in the CLT.

Proposition 3. Suppose that Assumptions 1, 3, 3′ and 4 hold. Then,

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

where $D_T = MV_TM'$ *and* $M = R[I, -Q_0Q^{-1}]$.

The asymptotic result in Proposition 3 shows that the asymptotic variance-covariance of the 2SQR(θ, q) estimator depends on the combination weight q through V_T, while the consistency of the slope estimator is not affected by the presence of q. To improve efficiency, q can be replaced with an optimal value (q^*) obtained by minimising the asymptotic covariance matrix shown in Proposition 3. However, there are many ways of minimising a multi-dimensional covariance matrix. For example, one may minimise some norm of the matrix (e.g., the mean square error). One may also minimise the standard error for a given coefficient of interest in the structural model.

For all these procedures and in some special cases (e.g., IID), the effect of q on D_T is concentrated in a scalar function that gathers the contribution of all the error terms to this matrix. Then, a unique and explicit solution q^* can be obtained. In the general case, q^* can also be made explicit when the MSE is minimised. Consistent preliminary estimators of q^* do not perturb the asymptotic properties of the 2SQR, which can be characterised as a MINPIN estimator (Andrews, 1994, p. 2263), as long as a stochastic equicontinuity condition of the global empirical process is valid.

We now exhibit a case with an explicit formula for q^*. Assume $\{(x_t', u_t, v_t)\}$ is iid and $f_t(0|x_t) = f(0)$, for any t. Then, the asymptotic covariance matrix in Proposition 3 simplifies into $\sigma_0^2(q)Q_{zz}^{-1}$, where $\sigma_0^2(q) = E(\zeta_t^2), \zeta_t = qf(0)^{-1}\psi_\theta(v_t) + u_t^* - qv_t^*$ and $Q_{zz} = H(\Pi_0^*)'QH(\Pi_0^*)$. In this case, we can easily obtain the optimal weight as in the following lemma.

Lemma 2. Suppose that Assumptions 1,3, 3′ and 4 hold. In addition, we assume that $\{(x_t', u_t, v_t)\}$ is iid and $f_t(0|x_t) = f(0)$, for any t. Then, the optimal weight minimizing the asymptotic variance of $\tilde{\alpha}$ is given by:

$$q^* = \frac{E(v_t^* u_t^*) - f(0)^{-1}E(\psi_\theta(v_t)u_t^*)}{f(0)^{-2}\theta(1-\theta) + E(v_t^{*2}) - 2f(0)^{-1}E(\psi_\theta(v_t)v_t^*)}. \tag{8}$$

A consistent estimator for q^* is obtained by substituting a consistent kernel-estimator $\hat{f}(0)$ for $f(0)$, and residuals for error terms:

$$\hat{q} = \frac{\sum_{t=1}^T \hat{v}_t^* \hat{u}_t^* - \hat{f}(0)^{-1}\sum_{t=1}^T \psi_\theta(\hat{v}_t)\hat{u}_t^*}{T\hat{f}(0)^{-2}\theta(1-\theta) + \sum_{t=1}^T \hat{v}_t^{*2} - 2\hat{f}(0)^{-1}\sum_{t=1}^T \psi_\theta(\hat{v}_t)\hat{v}_t^*}, \tag{9}$$

where $\hat{u}_t^* = \hat{v}_t^* - \hat{V}_t^{*\prime}\hat{\gamma}, \hat{v}_t^* = y_t - x_t'\tilde{\pi}, \hat{V}_t^{*\prime} = Y_t' - x_t'\tilde{\Pi}, \hat{v}_t = y_t - x_t'\hat{\pi}_\theta$ and $\hat{\pi}_\theta = \arg\min_\pi \sum_{t=1}^T \rho_\theta(y_t - x_t'\pi)$. The omitted proof for the consistency of $\hat{q}$ is straightforward.

To address robustness concerns, we now propose an estimator based on a robust first-stage estimator: the symmetrically trimmed-LS estimator (TLS). The TLS of π in the model $y = X\pi + v$ is $\hat{\pi}_{TLS} = (X'AX)^- XAy$, where $A = (a_{ij}), i, j = 1, ..., p$ and $a_{ij} = I_{[i=j \text{ and } X_i'\hat{\pi}(\mu) < y_i < X_i'\hat{\pi}(1-\mu)]}$, $\hat{\pi}(\mu)$ is the quantile regression estimator centered on a given quantile μ to be chosen a priori. Chen and Portnoy (1996) provide the TLS Bahadur representation. Let $\check{\alpha}$ be the estimator built from the TLS in the first stage and the quantile regression in the second stage. We adjust Assumptions 3′ and 4 as follows, with analogous interpretations of the different conditions.

Assumption 3″ (i) $E(v_t|x_t) - E(t\_v_t) - \mu\left[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)\right] = 0$,
where $t\_v_t \equiv v_t.I_{[F_v^{-1}(\mu) < v_t < F_v^{-1}(1-\mu)]}$ is the truncated error term.
(ii) $E(V_{it}|x_t) - E(t\_V_{it}) - \mu\left[F_{V_i}^{-1}(\mu) + F_{V_i}^{-1}(1-\mu)\right] = 0$, where V_{it} is the i^{th} element of V_t and $t\_V_{it} = V_t.I_{[F_V^{-1}(\mu) < V_t < F_V^{-1}(1-\mu)]}$ is the truncated error term.

Assumption 4′ (i) There exist finite constants $\Delta_{\tilde{v}}$ and $\Delta_{\bar{V}_j}$ such that $E|x_{ti}\tilde{v}_t^*|^3 < \Delta_{\tilde{v}}$ and

$E|x_{ti}\tilde{V}^*_{jt}|^3 < \Delta_{\bar{V}_j}$, for all i, j and t, where

$$\begin{aligned} \tilde{v}^*_t &= v_t - E(t\_v_t) - \mu\left[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)\right] \\ \text{and } \tilde{V}^*_{jt} &= V_{jt} - E(t\_V_{it}) - \mu\left[F_{V_j}^{-1}(\mu) + F_{V_j}^{-1}(1-\mu)\right]. \end{aligned}$$

(ii) The covariance matrix $\tilde{V}_T = var\left(T^{-1/2}\sum_{t=1}^T \tilde{S}_t\right)$ is positive definite for T sufficiently large, where $\tilde{S}_t = (q\psi_\theta(v_t), q\tilde{v}^*_t - \tilde{u}^*_t)' \otimes x_t$.

The Bahadur representation, obtained from Proposition 1, and the representation for TLS (Ruppert and Carroll, 1980) can be used to obtain the following asymptotic representation of $\check{\alpha}$:

$$\begin{aligned} T^{1/2}(\check{\alpha} - \alpha_0 - \tilde{B}_\alpha) &= RT^{-1/2}\sum_{t=1}^T x_t q\psi_\theta(v_t) \\ &\quad -RQ_0Q^{-1}T^{-1/2}\sum_{t=1}^T x_t(q\tilde{v}^*_t - \tilde{u}^*_t) + o_p(1), \end{aligned} \tag{10}$$

where the bias vector $\tilde{B}_\alpha$ is such that the first element only is non-zero. The non-zero bias term is given by $(1-q)\{E(t\_v_t) + \mu\left[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)\right]\} - \tilde{B}_{\Pi,1}\gamma_0$, where $\tilde{B}_{\Pi,1}$ is a G-row vector whose j^{th} element is given by $E(t\_V_{jt}) + \mu\left[F_{V_j}^{-1}(\mu) + F_{V_j}^{-1}(1-\mu)\right]$. The asymptotic representation (10) together with Assumptions $3''$ and $4'$ delivers the following asymptotic normality of $\check{\alpha}$.

Proposition 4. Suppose that Assumptions 1,3, $3''$ and $4'$ hold. Then,

$$\tilde{D}_T^{-1/2}T^{1/2}(\check{\alpha} - \alpha_0 - \tilde{B}_\alpha) \xrightarrow{d} N(0, I),$$

where $\tilde{D}_T = M\tilde{V}_T M'$ *and* $M = R[I, -Q_0Q^{-1}]$.

As before, if $\{(x_t', u_t, v_t)\}$ is iid and $f_t(0|x_t) = f(0)$ for any t, then the asymptotic matrix of $\check{\alpha}$ is $\tilde{\sigma}_0^2(q)\tilde{Q}_{zz}^{-1}$, where $\tilde{\sigma}_0^2(q) = E(\tilde{\zeta}_t^2), \tilde{\zeta}_t = qf(0)^{-1}\psi_\theta(v_t) + \tilde{u}^*_t - q\tilde{v}^*_t$ and $\tilde{Q}_{zz} = H(\tilde{\Pi}^*_0)'Q_0H(\tilde{\Pi}^*_0)$, and $\tilde{\Pi}^*_0 = \Pi_0 + \tilde{B}_\Pi$. In this case, it can be proved that the optimal value of q minimizing $\tilde{\sigma}_0^2(q)$ is given by:

$$q^* = \frac{E(\tilde{v}^*_t\tilde{u}^*_t) - f(0)^{-1}E(\psi_\theta(v_t)\tilde{u}^*_t)}{f(0)^{-2}\theta(1-\theta) + E(\tilde{v}^{*2}_t) - 2f(0)^{-1}E(\psi_\theta(v_t)\tilde{v}^*_t)}. \tag{11}$$

A consistent estimator for q^* is similarly obtained as follows:

$$\hat{q} = \frac{\sum_{t=1}^T \check{v}^*_t\check{u}^*_t - \hat{f}(0)^{-1}\sum_{t=1}^T \psi_\theta(\hat{v}_t)\check{u}^*_t}{T\hat{f}(0)^{-2}\theta(1-\theta) + \sum_{t=1}^T \check{v}^{*2}_t - 2\hat{f}(0)^{-1}\sum_{t=1}^T \psi_\theta(\hat{v}_t)\check{v}^*_t}, \tag{12}$$

where $\check{u}^*_t = \check{v}^*_t - \check{V}^{*'}_t\check{\gamma}$, $\check{v}^*_t = y_t - x_t'\hat{\pi}_{TLS}$, $\check{V}^{*'}_t = Y_t' - x_t'\hat{\Pi}_{TLS}$, $\hat{v}_t = y_t - x_t'\hat{\pi}_\theta$ and $\hat{\pi}_\theta = \arg\min_\pi \sum_{t=1}^T \rho_\theta(y_t - x_t'\pi)$.

In the next section, we present Monte Carlo simulation results, notably showing how much variance reduction can be achieved in finite samples with our method.[9]

[9] The case where the first stage is a quantile regression with the same quantile as in the second stage is reported in Kim and Muller (2004), with directly comparable tables.

6 Monte Carlo Simulations

6.1 Simulation Set-up

The data generating process used in the simulations is described in Appendix B. We study the finite sample properties of our two proposed two-stage estimators: (1) the OLS plus quantile regression estimator (2SQR1), and (2) the TLS plus quantile regression estimator (2SQR2). We impose $E(\psi_\theta(v_t)|x_t) = 0$ for each given θ. That is: for each θ, we regenerate the error terms such that $E(\psi_\theta(v_t)|x_t) = 0$ is satisfied, which means that we consider models centered differently according to the different θ. As explained in Appendix B, the equation of interest is assumed to be over-identified and the parameter values are set to $\beta' = (\beta_{0,1}, \beta_{0,2}) = (1, 0.2)$ and $\gamma = 0.5$. We generate the error terms by using three alternative distributions: the standard normal N(0,1), the Student-t with 3 degrees of freedom $t(3)$ and the Lognormal LN(0,1). The exogenous variables x_t are drawn independently from the errors from a normal distribution. For each of the 1000 replications, we estimate the parameter values β and γ using 2SQR1 and 2SQR2, and we calculate the deviations of the estimates from the true values. Then, we display the sample mean and sample standard deviation of these deviations over the 1,000 replications. The optimal value q^* is obtained by simulating the formula in (8) or (11), while $\hat{q}$ is estimated through (9) or (12).

6.2 Results

We first discuss the results for the 2SQR1(θ, q) with N(0,1), $t(3)$ and LN(0,1) errors, shown in Tables 1-3 for the case of iid errors.[10] We first discuss the results when the error terms are drawn from N(0,1) shown in Table 1. In all cases, as expected, the intercept estimate exhibits biases that do not vanish as the sample size increases. On the other hand, the 2SQR1(θ, q) estimates for the slope parameters (β_1 and γ) are unbiased for all specifications, all evaluations of q and all $\theta's$ and even with a sample size as small as 50. Using the optimal value q^* dramatically improves the accuracy of the 2SQR1(θ, q) as compared to the case $q = 1$. The optimal values q^* are close to zero, which can be viewed as related to a kind of inverse least-squares extraction of the structural parameters from the reduced-form parameters. This is what pushing $\hat{q}$ to zero does, as can be seen in (5). The gain is larger for the extreme quantiles ($\theta = 0.05$ and 0.95) than for the middle quantiles ($\theta = 0.25, 0.5$ and 0.75). Even with $T = 50$, using $\hat{q}$ can substantially improve efficiency as compared to $q = 1$. The estimation accuracy of $\hat{q}$ and the efficiency gain improve as the sample size increases. With $T = 300$, using $\hat{q}$ or q^* is almost indifferent for estimating α_0, although the estimated values of $\hat{q}$ are not always very close to q^*.

Table 2 shows the results for the Student-t distribution case. As expected, with fat tails $t(3)$ errors the standard deviations of the sampling distributions of the 2SQR1(θ, q) are much larger than with normal errors. As before, the variance reductions from using q^* are small for middle quantiles, while substantial reductions can be achieved for extreme quantiles. The standard deviations are the largest for the lognormal case, where using q^* always yields outstanding efficiency gains. For right-skewed distributions, quantile regressions are typically inaccurate for large quantiles. In this case, our method generates large efficiency gains. For example, considering the case with $T = 300$, the standard error for $\hat{\gamma}$ with $q = 1$ is 0.91, while it is reduced to 0.25 with $q = \hat{q}$. However, there is virtually no efficiency gain with small values of θ less than around 0.5.

Given the generally substantial efficiency gains, it is natural to ask how close the reduced variance is to the Cramer-Rao lower bound. We have calculated the CR bound numerically for

[10] We have conducted the same set of simulations for the case of heteroscedastic errors and have found that the results are qualitatively the same as in the iid case.

each distribution in Table 4(a) for $T = 50$ and Table 5(a) $T = 300$. Table 4(a) shows the simulated asymptotic standard deviations for 2SLS and 2SQR1$(\theta, \hat{q})$ with $\theta = 0.25, 0.50, 0.95$, along with the simulated CR bounds with $T = 50$. We only discuss the slope coefficients as the intercept coefficient estimate is biased.

For a small sample size such as $T = 50$, the 2SLS efficiency loss is not negligible even for the normal distribution case and the efficiency loss becomes surprisingly large for both $t(3)$ and LN(0,1). On the other hand, 2SQR1$(\theta, \hat{q})$ attains the CR bounds at the middle quantiles such as $\theta = 0.25, 0.5$ and 0.75 for the normal distribution case. Moving to the Student-t distribution case, 2SQR1$(\theta, \hat{q})$ is much more efficient than the 2SLS again at the middle quantiles ($\theta = 0.25, 0.5$ and 0.75). For example, we observe approximately 24% efficiency gain at $\theta = 0.5$. Finally, under lognormality, 2SLS performs badly relative to the CR bounds, while 2SQR1$(\theta, \hat{q})$ stays closer to the CR bounds for small and middle quantiles than 2SLS.

When we increase the sample size to $T = 300$, we have qualitatively the same results as shown in Table 5(a). For the normal errors, both 2SLS and 2SQR1$(\theta, \hat{q})$ attain the CR bounds. However, all the previous observations still hold for the other two error distributions t(3) and LN(0,1); i.e., (i) 2SQR1$(\theta, \hat{q})$ is more efficient than 2SLS at the middle quantiles for $t(3)$, and (ii) 2SQR1$(\theta, \hat{q})$ is more efficient than 2SLS at the low and middle quantiles for LN(0,1).

Let us now turn to 2SQR2 based on the TLS at the first stage.[11] According to our simulations, using the trimming value of $\mu = 0.25$ yields more accurate results than other trimming values such as 0.05 or 0.10.[12] For a large sample size such as $T = 300$, trimming at $0.05, 0.10$ or 0.25 is almost indifferent. Hence, we focus on the case $\mu = 0.25$. The results for 2SQR2 are reported in Tables 4(b) and 5(b), respectively for $T = 50$ and $T = 300$ where the two cases ($q = 1$ and $q = \hat{q}$) are shown and compared. As clearly demonstrated in the two tables, our proposed variance reduction method works again very well with the 2SQR2 estimator.

The 2SQR2$(\theta, \hat{q})$ appears to perform uniformly better than the 2SQR2$(\theta, q = 1)$, except for $T = 50$ at the median for $t(3)$ and at a few low quantiles for LN(0,1) – probably because of sampling errors since this irregularity vanishes when $T = 300$. The improvement from moving from $q = 1$ to $q = \hat{q}$ is sizeable at quantile 0.95 for symmetric errors (up to 60% reduction in standard deviation) and at large quantiles for asymmetric errors (up to 80% reduction). The 2SQR2$(\theta, \hat{q})$ clearly improves on the 2SQR1$(\theta, \hat{q})$ for both $t(3)$ and LN(0,1), while the reverse is true for normal errors.

Let us finally compare the most interesting two estimators; i.e. 2SQR1$(\theta, \hat{q})$ and 2SQR2$(\theta, \hat{q})$ in Tables 4 and 5. Under normal errors, the 2SQR1$(\theta, \hat{q})$ and the 2SQR2$(\theta, \hat{q})$ both almost reach the CR bound, whatever the considered quantile. Here, reformulating the dependent variable is fruitful, especially for upper quantiles for which it allows massive efficiency gains. The 2SQR2$(\theta, q = 1)$ is slightly outperformed by the 2SQR1$(\theta, \hat{q})$, perhaps because trimming here only discards information**.** With Student errors, the 2SQR2$(\theta, \hat{q})$ is often the more accurate estimator, yielding results fairly close to the CR bound. Under lognormality, none of the studied estimators approaches the CR bound. However, using the 2SQR2$(\theta, \hat{q})$ generally yields the best accuracy. For upper quantiles, reformulating the dependent variables delivers huge efficiency gains.

It is interesting to reflect on the proximity of the results of the 2SQR1$(\theta, \hat{q})$ and the 2SQR2$(\theta, 1)$ in the light of the non-robustness of the OLS and the robustness of the TLS. Redefining the dependent variable may improve the robustness of the two-stage estimator through the reduction

[11] We have also tried LAD in the first-stage, as in Chen and Portnoy (1996). However, the results are almost identical to that of the 2SQR2 so that we do not include the results in the paper, while they are available upon request. What seems to matter here is the robustness of the first stage estimator.

[12] One exception is the nomal distribution case with $T = 50$, in which case trimming at 0.25 is only slightly inferior than 0.1.

of the influence of outliers for the errors v_t, even when the first-stage estimator is non-robust. This effect, apparent in the formula of the asymptotic representation, is confirmed in the small sample simulations. Thus, specific estimators of q could also be chosen to enhance robustness.

7 Conclusion

In this paper, we develop a new method of variance reduction for two-stage estimation procedures and apply it to the case of two-stage quantile regression allowing for random regressors as well as non-iid error terms. In this setting, we show that an asymptotic bias that would occur in the first-stage reduced-form estimates of the coefficients of the exogenous variables in the structural equation is integrally and exclusively transmitted to the coefficients of the same variables in the second-stage. At this occasion, we show that the structural approach and the fitted-value approach to two-stage quantile regressions amount to estimating the same structural slope coefficients, under a natural instrumental variable assumption corresponding to the constant effects case of quantile regressions.

Following an original idea in Amemiya (1982), we reformulate the dependent variable as a weighted mean of the original dependent variable and its fitted value. Such a combination introduces a trade-off between an asymptotic bias on the intercept of the equation of interest on the one hand, and the variance reduction of the slope estimator on the other hand. Using such a trade-off, we can improve the efficiency of the slope estimator at the expense of making the intercept estimator inconsistent.

We derive the asymptotic normality and the asymptotic variance-covariance matrix of such two-stage quantile regression estimators. Then, we apply our variance reduction method to these specific two-stage quantile regression estimators. Our Monte Carlo simulation results show massive efficiency gains in many cases. In particular, our new method alleviates the well-known poor efficiency of quantile regressions at extreme quantiles. Two important arguments to use quantile regression jointly with variance-reduction are first that it yields more accurate or equivalent estimates than OLS, and second that it does not require the knowledge of the distribution shape, which is a drawback of maximum likelihood estimators.

Let us emphasize two practical principles in our approach. First, the first-stage estimators should be carefully selected so as to preserve efficiency, robustness or other desired properties. Our simulation results suggest that OLS should perform well under normality, while trimmed least-square should be more accurate and more robust for heavy tails or asymmetric error distributions. Second, one should reformulate the dependent variable as proposed, in such a way that a selected variance criterion is minimized. The choice of the variance criterion may be left to the researcher, while minimising the MSE seems to be a natural choice.

We finally recap the computation steps for trimmed least-squares plus quantile regression: (1) trimmed least-squares for the reduced-form equation and the ancillary equations, (2) computation of the fitted-values for the endogenous regressors of the structural equation, (3) preliminary quantile regression of the structural equation where the endogenous regressors are substituted with their fitted values, (4) estimation of the density of the reduced form error at the quantile of interest, (5) estimation of the optimal weight for the reformulation, using residuals and density estimates from the previous stages, (6) reformulation of the dependent variable using the optimal weight, (7) final quantile regression of the structural equation.

It is worth mentioning as a word of conclusion that the proposed method for variance reduction is not necessarily limited to two-stage quantile regression. In fact, our approach might be generalized to any two-stage regression where the use of a composite dependent variable does not disturb the

consistency property of the final estimator.

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, 91-117.

[2] Amemiya, T. (1982), "Two Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689-711.

[3] Andrews, D.W.K. (1990), "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity and Nonparametric Kernel Estimation," Discussion paper at the Cowles Foundation for Research in Economics, Yale University.

[4] Andrews, D.W.K. (1994), "Empirical Process Methods in Econometrics," *Handbook of Econometrics*, Volume IV, eds. R.F. Engle and D.L. McFadden, New York: North-Holland, pp. 2247-94.

[5] 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.

[6] Chen, L.-A. (1988), "Regression Quantiles and Trimmed Least-Squares Estimators for Structural Equations and Non-Linear Regression Models," Unpublished Ph.D. dissertation, University of Illinois at Urbana-Champaign.

[7] 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, 1005-1032.

[8] Chen, X., O. Linton and I. Van Keilegem (2003), "Estimation of Semi-Parametric Models When the Criterion Function is not Smooth," *Econometrica*, 71, 1591-1608.

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

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

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

[12] Chernozhukov, V. and C. Hansen (2008), "The Reduced Form: A Simple Approach to Inference with Weak Instruments," Economics Letters, 100, 68-71.

[13] Chernozhukov, V., G.W. Imbens and W.K. Newey (2007), "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4-14.

[14] Chesher, A. (2003), "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.

[15] Chevapatrakul, T., T. Kim and P. Mizen (2009), "The Taylor Principle and Monetary Policy Approaching a Zero Bound on Nominal Rates: Quantile Regression Results for the United States and Japan," *Journal of Money, Credit and Banking*, 41, 1705-1723.

[16] Chortareas, G., G. Magonis and T. Panagiotidis (2012), "The Asymmetry of the New Keynesian Phillips Curve in the Euro Area," *Economics Letters*, 114, 161-163.

[17] Garcia, J., P.J. Hernandez and A. Lopez (2001), "How Wide is the Gap? An Investigation of Gender Wage Differences Using Quantile Regression," *Empirical Economics*, 26, 149-167.

[18] Hjort, N.L. and D. Pollard (1999), "Asymptotics for Minimisers of Convex Processes," Discussion paper, Yale University.

[19] Hong, H. and E. Tamer (2003), "Inference in Censored Models with Endogenous Regressors," *Econometrica*, 71, 905-932.

[20] Honore, B.E. and L. Hu (2004), "On the Performance of Some Robust Instrumental Variables Estimators," *Journal of Business and Economic Statistics*, 22, 30-39.

[21] Imbens, G.W. and W.K. Newey (2006), "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," Discussion paper, MIT Department of Economics.

[22] James, W. and C. Stein (1960), "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (vol. 1), Berkeley, CA: University of California Press, 361-379.

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

[24] Kemp, G.C.R. (1999), "Least Absolute Error Difference Estimation of a Single Equation from a Simultaneous Equations System," Discussion paper, University of Essex, December.

[25] Kim, T. and C. Muller (2004), "Two-Stage Quantile Regressions when the First Stage is Based on Quantile Regressions," *The Econometrics Journal*, 18-46.

[26] Kim, T. and H. White (2001), "James-Stein-Type Estimators in Large Samples with Application to the Least Absolute Deviations Estimator," *Journal of the American Statistical Association*, 96, 697-705.

[27] Kim, T. and H. White (2003), "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," *Advances in Econometrics*, 17, 107-132.

[28] Koenker, R. (2005), "*Quantile Regression*," Cambridge University Press, New York.

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

[30] Koenker, R. and Q. Zhao (1996), "Conditional Quantile Estimation and Inference for ARCH models," *Econometric Theory*, 12, 793-813.

[31] Lee, S. (2007), "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, 141, 1131-1158.

[32] Ma, L. and R. Koenker (2006), "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics,* 134, 471-506.

[33] MaCurdy, T. and C. Timmins (2000), "Bounding the Influence of Attrition on Intertemporal Wage Variation in the NLSY," mimeo Stanford University, May.

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

[35] Pollard, D. (1991), "Asymptotic for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186-199.

[36] Portnoy, S. (1991), "Asymptotic Behaviour of Regression Quantiles in Non-Stationary, Dependent Cases," *Journal of Multivariate Analysis*, 38, 100-113.

[37] Powell, J. (1983), "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569-1575.

[38] Ruppert, D. and R.J. Carroll (1980), "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828-838.

[39] Sakata, S. (2007), "Instrumental Variable Estimation Based on Conditional Median Restriction," *Journal of Econometrics*, 141, 350-382.

[40] Sen, P.K. and A.K.M.E. Saleh (1987), "On Preliminary Test and Shrinkage M-estimation in Linear Models," *The Annals of Statistics*, 15, 1580-1592.

[41] Weiss, A.A. (1990), "Least Absolute Error Estimation in the Presence of Serial Correlation," *Journal of Econometrics*, 44, 127-158.

[42] White, H. (2001), *Asymptotic Theory for Econometricians*, Academic Press, San Diego.

[43] White, H., T. Kim, and S. Manganelli (2009), "Modeling Autoregressive Conditional Skewness and Kurtosis with Multi-Quantile CAViaR," "*Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle.*"

Appendix A: Mathematical Proofs

Proof of Lemma 1: Let $M_{Ti}^*(\zeta) = T^{-1/2}\sum_{t=1}^{T} m_i^*(w_t,\zeta)$, where ζ is a $K\times 1$ vector, $w_t = (v_t, x_t')'$, $m_i^*(w_t,\zeta) = x_{ti}\psi_\theta(qv_t - x_t'\zeta)$ and x_{ti} is the i^{th} element in x_t. We define $V_{Ti}^*(\zeta) = M_{Ti}^*(\zeta) - E(M_{Ti}^*(\zeta))$. We shall show that $\{V_{Ti}^*(\zeta) : T \geq 1\}$ is stochastically equicontinuous. To do so, we use Theorem II.8 in Andrews (1990) for which the following two conditions must be verified; (a) $m_i^*(w_t,\zeta)$ is a type IV class function with index $p \geq 2$; that is, for all bounded ζ in R^K and for all $L_1 > 0$ in a neighborhood of zero,

$$\sup_{t\leq T, T>1}\left[E\left(\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m_i^*(w_t,\zeta_1) - m_i^*(w_t,\zeta)|^p\right)\right]^{1/p} \leq CL_1^{\psi} \tag{13}$$

for some positive constants C and ψ and (b) $\{w_t\}$ is $\alpha-$mixing of size $-\frac{(2K+\psi)(K+2\psi)}{\psi^2}$.

We first verify (a) for $p = 2$. Consider a constant L_1 close to zero and a finite value of ζ in R^K. Note that

$$\begin{aligned} & |m_i^*(w_t,\zeta_1) - m_i^*(w_t,\zeta)| = |x_{ti}||1_{[qv_t - x_t'\zeta\leq 0]} - 1_{[qv_t - x_t'\zeta_1\leq 0]}| \\ \equiv\ & |x_{ti}||1_{[A\leq 0]} - 1_{[B\leq 0]}| \leq |x_{ti}||1_{[|A|\leq|A-B|]}| \leq |x_{ti}||1_{[|qv_t - x_t'\zeta|\leq||x_t||\times||\zeta_1-\zeta||]}|, \end{aligned}$$

where $A = qv_t - x_t'\zeta$ and $B = qv_t - x_t'\zeta_1$. Hence, we have

$$\begin{aligned} & \sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m_i^*(w_t,\zeta_1) - m_i^*(w_t,\zeta)|^2 \\ \leq\ & x_{ti}^2 \sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1} 1_{[|qv_t - x_t'\zeta|\leq||x_t||\times||\zeta_1-\zeta||]} \leq x_{ti}^2 1_{[|qv_t - x_t'\zeta|\leq||x_t||L_1]}, \end{aligned}$$

which implies

$$\begin{aligned} & E\left(\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m_i^*(w_t,\zeta_1) - m_i^*(w_t,\zeta)|^2\right) \\ \leq\ & E\left(x_{ti}^2 P_{x_t}\left[|qv_t - x_t'\zeta| \leq ||x_t||L_1\right]\right) = E\left(x_{ti}^2\int_{L_0}^{U_0} f_{v|x}(\lambda|x_t)d\lambda\right) \qquad (\because q > 0) \\ \leq\ & E\left(x_{ti}^2\int_{L_0}^{U_0} f_0 d\lambda\right) \qquad (\because \text{Assumption 3(ii)}) \\ =\ & \frac{2f_0}{q}E\left(x_{ti}^2||x_t||\right)L_1, \end{aligned}$$

where P_{x_t} is the conditional probability function given x_t, $U_0 = q^{-1}(x_t'\zeta + ||x_t||L_1)$ and $L_0 \equiv q^{-1}(x_t'\zeta - ||x_t||L_1)$. Hence,

$$\sup_{t\leq T, T>1}\left[E\left(\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m_i^*(w_t,\zeta_1) - m_i^*(w_t,\zeta)|^2\right)\right]^{1/2} \leq CL_1^{1/2}$$

for some constant C because of Assumption 3(v). Hence, condition (a) is satisfied with $\psi = 1/2$.

Next, we turn to condition (b). Since $\psi = 1/2$,

$$-\frac{(2K+\psi)(K+2\psi)}{\psi^2} = -\frac{(2K+\frac{1}{2})(K+1)}{\frac{1}{4}} = -2\left(4K+1\right)\left(K+1\right).$$

Hence, condition (b) is a consequence of Assumption 1. Therefore, by Theorem II.8 in Andrews (1990), $\{V_{Ti}^*(\zeta) : T \geq 1\}$ is stochastically equicontinuous, which implies that $V_T^*(\zeta)$ is also stochastically equicontinuous. Then, for any constant sequence L_T^* that converges to zero, we have

$$\sup_{||\zeta_1-\zeta_2||\leq L_T^*} ||V_T^*(\zeta_1) - V_T^*(\zeta_2)|| = o_p(1). \tag{14}$$

We now introduce a factor $T^{-1/2}$ that weighs the contribution of the first-stage estimator in the kernel of the empirical process. For this, we choose $L_T^* = T^{1/2}L$ for a fixed positive number L. Let $V_T(\Delta) = M_T(\Delta) - E(M_T(\Delta))$, where $M_T(\Delta) = T^{-1/2}\sum_{t=1}^T m(w_t, \Delta)$, $m(w_t, \Delta) = x_t\psi_\theta(qv_t - T^{-1/2}x_t'\Delta)$, and Δ is a $K \times 1$ vector. Since $V_T^*(\zeta) = V_T(T^{1/2}\zeta)$, by defining $\Delta_1 = T^{1/2}\zeta_1$ and $\Delta_2 = T^{1/2}\zeta_2$, the result in (14) becomes

$$\sup_{||\Delta_1-\Delta_2||\leq L} ||V_T(\Delta_1) - V_T(\Delta_2)|| = o_p(1). \tag{15}$$

Setting $\Delta_1 = \Delta$ and $\Delta_2 = 0$ in (15), yields

$$\sup_{\|\Delta\|<L} \|M_T(\Delta) - M_T(0) - \{EM_T(\Delta) - EM_T(0)\}\| = o_p(1). \tag{16}$$

Next, we show that $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$ as follows. First, we note that

$$E(M_T(\Delta)) = E\left\{T^{-1/2}\sum_{t=1}^T\left[x_t\theta - x_t\int_{-\infty}^{q^{-1}x_t'T^{-1/2}\Delta} f_t(v|x_t)dv\right]\right\}.$$

Therefore, we have

$$\begin{aligned} E(M_T(\Delta)) - E(M_T(0)) &= -E\left\{T^{-1/2}\sum_{t=1}^T\left[x_t\int_0^{q^{-1}x_t'T^{-1/2}\Delta} f_t(v|x_t)dv\right]\right\} \\ &= -E\left\{q^{-1}T^{-1}\sum_{t=1}^T x_tx_t'\Delta\frac{F_t(q^{-1}x_t'T^{-1/2}\Delta|x_t) - F_t(0|x_t)}{q^{-1}x_t'T^{-1/2}\Delta}\right\}, \end{aligned}$$

where $F_t(\cdot|x_t)$ is the conditional cdf of v_t. Let $G(\lambda) = q^{-1}T^{-1}\sum_{t=1}^T F_t(\lambda|x_t)x_tx_t'\Delta$. Then, by the Mean-Value Theorem and the continuity in Assumption 3(ii), there exists $\xi_{T,t}$ between 0 and $q^{-1}x_t'T^{-1/2}\Delta$ such that $E(M_T(\Delta))-E(M_T(0)) = -E\{G'(\xi_{T,t})\} = -q^{-1}E\{T^{-1}\sum_{t=1}^T f_t(\xi_{T,t}|x_t)x_tx_t'\}\Delta$. We now examine the convergence of this term.

Let $Q_T = E\left[T^{-1}\sum_{t=1}^T f_t(\xi_{T,t}|x_t)x_tx_t'\right]$, $Q_{0T} = E\left[T^{-1}\sum_{t=1}^T f_t(0|x_t)x_tx_t'\right]$ and consider the $(i,j)^{th}$ element of $|Q_T - Q_{0T}|$, which is given by

$$\begin{aligned} & |T^{-1}\sum_{t=1}^T E\left(\{f_t(\xi_{T,t}|x_t) - f_t(0|x_t)\}x_{ti}x_{tj}\right)| \\ \leq\ & T^{-1}\sum_{t=1}^T E\left(|f_t(\xi_{T,t}|x_t) - f_t(0|x_t)|\ |x_{ti}|\ |x_{tj}|\right) \\ \leq\ & L_0T^{-1}\sum_{t=1}^T E\left(|\xi_{T,t}|\ |x_{ti}|\ |x_{tj}|\right) \end{aligned}$$

for some constant L_0, where the first result is due to Minkowski's inequality and Jensen's inequality and the second result is obtained by the Lipschitz continuity in Assumption 3(ii). Next, we note that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} E\left(|\xi_{T,t}|\ |x_{ti}|\ |x_{tj}|\right) &\leq q^{-1}T^{-3/2}\sum_{t=1}^{T} E\left(\left|x_t'\Delta\right|\ |x_{ti}|\ |x_{tj}|\right) \\ &\leq q^{-1}\left\|\Delta\right\| T^{-3/2}\sum_{t=1}^{T} E\left(\left\|x_t\right\|^3\right) \\ &\leq q^{-1}\left\|\Delta\right\| T^{-1/2}C \to 0 \end{aligned}$$

for a constant C, where the last inequality is obtained by Assumption 3(v). Since $Q_0 = \lim_{T\to\infty} Q_{0T}$, we have $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$. QED.

Proof of Proposition 1: We define $\hat{\Delta}_0 = -(1-q)T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) + T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0$. We have $\hat{\Delta}_0 = O_p(1)$ because of Assumption 2. Then, Lemma 1 implies that

$$M_T(\hat{\Delta}_0) = M_T(0) - q^{-1}Q_0\hat{\Delta}_0 + o_p(1), \tag{17}$$

where M_T is defined Lemma 1. The term $q^{-1}Q_0\hat{\Delta}_0$ is bounded in probability because $\hat{\Delta}_0 = O_p(1)$. Also, $M_T(0) = T^{-1/2}\sum_{t=1}^{T} x_t\psi_\theta(qv_t) = T^{-1/2}\sum_{t=1}^{T} x_t\psi_\theta(v_t)$ because $q > 0$.[13] Therefore, under Assumptions 1, 3(iv)-(v) and 4(i), $T^{-1/2}\sum_{t=1}^{T} x_t\psi_\theta(v_t)$ converges in distribution to a normal random variable by the CLT in Theorem 5.20 of White (2001). Therefore, we have

$$M_T(\hat{\Delta}_0) = O_p(1). \tag{18}$$

Next, we define $\hat{\Delta}_1(\delta) = H(\hat{\Pi})\delta + \hat{\Delta}_0 = H(\hat{\Pi})\delta - (1-q)T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) + T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0$ for $||\delta|| \leq L$, where $\delta \in R^{G+K_1}$ for some $L > 0$. Using Assumption 2 and Lemma 1, it is straightforward to show that

$$\sup_{||\delta||\leq L} ||M_T(\hat{\Delta}_1(\delta)) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta)|| = o_p(1). \tag{19}$$

Before to reach the main part of the proof, we need one more result of stochastic equicontinuity. For this, we define $\tilde{M}_T(\delta) = H(\hat{\Pi})'M_T(\hat{\Delta}_1(\delta))$ and $||H(\hat{\Pi})||^2 = tr(H(\hat{\Pi})H(\hat{\Pi})')$, which is $O_p(1)$ since $\hat{\Pi}$ converges to $\Pi_0 + B_\Pi$ that is finite.

We now use the argument between (A.7) and (A.8) in Powell (1983) to show that (18) and (19) imply that for some finite $L_2 > 0$:

$$\sup_{||\delta||\leq L_2} ||\tilde{M}_T(\delta) - H(\Pi_0^*)'M_T(\hat{\Delta}_0) + q^{-1}Q_{zz}^*\delta|| = o_p(1), \tag{20}$$

where $Q_{zz}^* = H(\Pi_0^*)'Q_0H(\Pi_0^*)$. The essence of the Powell's argument is the following. Since $||H(\hat{\Pi})||^2 = O_p(1)$ and $||H(\hat{\Pi}) - H(\Pi_0^*)|| = o_p(1)$, we have

$$\begin{aligned} & ||\tilde{M}_T(\delta) - H(\Pi_0^*)'M_T(\hat{\Delta}_0) + Q_{zz}^*\delta|| \\ \leq\ & ||H(\hat{\Pi})||\ ||M_T(\hat{\Delta}_1(\delta)) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta)|| + ||H(\hat{\Pi}) - H(\Pi_0^*)||\ ||M_T(0) - q^{-1}Q_0\hat{\Delta}_0|| \\ & + ||H(\hat{\Pi}) - H(\Pi_0^*)||\ \left\{||H(\hat{\Pi})||\ + ||H(\Pi_0^*)||\right\} ||q^{-1}Q_0||||\delta|| + ||H(\Pi_0^*)||, \end{aligned}$$

[13] For $q < 0$, we have $\psi_\theta(qv_t) = -\psi_{1-\theta}(v_t)$. Therefore, $E(\psi_\theta(v_t)|x_t) = 0$ does not imply $E(\psi_\theta(qv_t)|x_t) = 0$ in general, except for LAD estimators ($\theta = 1/2$) or symmetric distributions. This might be one reason why authors imposed symmetry of error terms, as in Chen (1988) and Chen and Portnoy (1996).

which delivers the result by applying the sup-operator to both sides of the inequality above.

Next, we define $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$, where the expression for B_α is given in the proposition. We wish to show that

$$\tilde{M}_T(\hat{\Delta}) = o_p(1). \tag{21}$$

Note that

$$\begin{aligned}
\tilde{M}_T(\hat{\Delta}) &= H(\hat{\Pi})' M_T(\hat{\Delta}_1(\hat{\Delta})) \\
&= T^{-1/2} \sum_{t=1}^{T} H(\hat{\Pi})' x_t \psi_\theta(q v_t - T^{-1/2} x_t' \hat{\Delta}_1(\hat{\Delta})) \\
&= T^{-1/2} \sum_{t=1}^{T} H(\hat{\Pi})' x_t \psi_\theta(q y_t + (1-q)\hat{y}_t - x_t' H(\hat{\Pi})\hat{\alpha} + \hat{A}_t + \hat{B}_t),
\end{aligned}$$

where

$$\begin{aligned}
\hat{A}_t &= x_t' H(\hat{\Pi})\alpha_0 - x_t' H(\Pi_0)\alpha_0 + x_t'\Pi_0\gamma_0 - x_t'\hat{\Pi}\gamma_0 \\
\text{and } \hat{B}_t &= x_t'[H(\hat{\Pi})B_\alpha - (1-q)B_\pi + B_\Pi\gamma_0].
\end{aligned}$$

First, we have that $\hat{A}_t = 0$ because $x_t' H(\hat{\Pi})\alpha_0 = x_{1t}'\beta_0 + x_t'\hat{\Pi}\gamma_0$ and $x_t' H(\Pi_0)\alpha_0 = x_{1t}'\beta_0 + x_t'\Pi_0\gamma_0$. Moreover, $\hat{B}_t = 0$ because of the definition of B_α. Since $\hat{A}_t = 0$ and $\hat{B}_t = 0$, it can be shown that $T^{1/2}\tilde{M}_T(\hat{\Delta}) = \left[\frac{\partial S_T}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}} \right]_-$, which is the vector of left-hand-side partial derivatives of the objective function in (5) evaluated at the solution $\hat{\alpha}$. Therefore, we obtain the desired result in (21); i.e., $\tilde{M}_T(\hat{\Delta}) = o_p(1)$.

Next, let us show that $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1)$. This will prove that B_α is the asymptotic bias of $\hat{\alpha}$. We can obtain $\hat{\Delta} = O_p(1)$ by using the argument in Lemma A.4 in Koenker and Zhao (1996). Similar arguments are in Jureckova (1977) and Hjort and Pollard (1999). To use Lemma A.4 in Koenker and Zhao (1996) and to obtain $\hat{\Delta} = O_p(1)$, we need to check the following conditions:
(i) $-\delta'\tilde{M}_T(\lambda\delta) \geq -\delta'\tilde{M}_T(\delta)$ for $\lambda \geq 1$ and $||\delta|| \leq L_3$ for some $L_3 > 0$,
(ii) $||H(\Pi_0^*)' M_T(\hat{\Delta}_0)|| = O_p(1)$,
(iii) $\tilde{M}_T(\hat{\Delta}) = o_p(1)$,
(iv) Q_{zz}^* is positive definite.
Condition (i) is obtained by noticing that function $h(\lambda) = \sum_{t=1}^{T} \rho_\theta(q v_t - T^{-1/2} x_t' H(\hat{\Pi})\delta\lambda - T^{-1/2} x_t'\hat{\Delta}_0)$ is convex in λ, and therefore that its gradient, $-\delta'\tilde{M}_T(\lambda\delta)$ is non-decreasing in λ. Condition (ii) comes from (18). Condition (iii) results from the first-order condition of the second stage, as we discussed above. Finally, condition (iv) is ensured by Assumptions 3(i) and 3(iii). Hence, by Lemma A.4 in Koenker and Zhao (1996), we have

$$\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1). \tag{22}$$

Therefore, we can plug $\hat{\Delta}$ into (20) in place of δ to obtain the following result:

$$\tilde{M}_T(\hat{\Delta}) - H(\Pi_0^*)' M_T(\hat{\Delta}_0) + q^{-1} Q_{zz}^* \hat{\Delta} = o_p(1). \tag{23}$$

Note that the first term in (23) is $o_p(1)$ because of (21). Hence, we have

$$\begin{aligned}
q^{-1} Q_{zz}^* \hat{\Delta} &= H(\Pi_0^*)' M_T(\hat{\Delta}_0) + o_p(1) &(24)\\
&= H(\Pi_0^*)' M_T(0) - H(\Pi_0^*)' q^{-1} Q_0 \hat{\Delta}_0 + o_p(1), &(25)
\end{aligned}$$

where the second equality comes from (17). By plugging the definition of $\hat{\Delta}_0$ and inverting $q^{-1}Q_{zz}^*$, we obtain

$$\begin{aligned} T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) &= Q_{zz}^{*-1} H(\Pi_0^*)' \{T^{-1/2} \sum_{t=1}^{T} x_t q \psi_\theta(v_t) \\ &\quad + (1-q) Q_0 T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) - Q_0 T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0\} + o_p(1), \end{aligned}$$

which completes the proof. QED.

Proof of Proposition 2: Let us decompose B_α as follows: $B_\alpha = \begin{bmatrix} B_{\alpha,1} \\ B_{\alpha,2} \end{bmatrix}$ where $B_{\alpha,1}$ is the vector of the K_1 elements of B_α. We need to show that (i) $B_{\alpha,1} = (1-q)B_{\pi,1} - B_{\Pi,1}\gamma_0$, and (ii) $B_{\alpha,2} = 0_G$. First, we note that the definition of B_α is given by $B_\alpha = RQ_0\{(1-q)B_\pi - B_\Pi\gamma_0\}$ in Proposition 1. Second, we decompose matrix R as $R = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix}$, where R_1 is the first $K_1 \times K$ submatrix and R_2 is the remaining $G \times K$ submatrix. We decompose matrix Q_0 as $Q_0 = [Q_1, Q_2]$ where Q_1 is the first $K \times K_1$ submatrix and Q_2 is the remaining $K \times K_2$ submatrix. Then, we have

$$B_\alpha = \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} [Q_1, Q_2] \begin{bmatrix} B_{\alpha,1} \\ B_{\alpha,2} \end{bmatrix} = \begin{bmatrix} R_1 Q_1 B_{\alpha,1} \\ R_2 Q_1 B_{\alpha,1} \end{bmatrix},$$

where the second equality comes from the fact that $B_{\alpha,2} = 0_G$. Therefore, if we can show that (i) $R_1 Q_1 = I_{K_1}$, and (ii) $R_2 Q_1 = 0_{G \times K_1}$, then the proof will be completed. Because of the definition of R, the following identity holds.

$$RQ_0 H(\Pi_0^*) = I_{K_1 \times G}. \tag{26}$$

Let us decompose Π_0^* as $\Pi_0^* = \begin{bmatrix} \Pi_{0,1}^* \\ \Pi_{0,2}^* \end{bmatrix}$, where $\Pi_{0,1}^*$ is the $K_1 \times G$ submatrix and $\Pi_{0,2}^*$ is the remaining $K_2 \times G$ submatrix. With this decomposition of Π_0^*, The left-hand-side of (26) is given by

$$RQ_0 H(\Pi_0^*) = \begin{bmatrix} R_1 Q_1 & R_1 Q_2 \\ R_2 Q_1 & R_2 Q_2 \end{bmatrix} \begin{bmatrix} I_{K_1} & \Pi_{0,1}^* \\ 0_{K_2 \times K_1} & \Pi_{0,2}^* \end{bmatrix} = \begin{bmatrix} R_1 Q_1 & R_1 Q_1 \Pi_{0,1}^* + R_1 Q_2 \Pi_{0,2}^* \\ R_2 Q_1 & R_2 Q_1 \Pi_{0,1}^* + R_2 Q_2 \Pi_{0,2}^* \end{bmatrix}.$$

Noting that the (1,1)-block is a $K_1 \times K_1$ submatrix and the (2,2)-block is a $G \times G$ submatrix, the identity in (26) delivers the desired results: (i) $R_1 Q_1 = I_{K_1}$, and (ii) $R_2 Q_1 = 0_{G \times K_1}$. QED.

Proof of Proposition 3: Replacing the asymptotic representation of the first-stage estimators and collecting terms in the asymptotic representation for the 2SQR(θ, q) with LS first-stage estimators gives

$$T^{1/2}(\tilde{\alpha} - \alpha_0 - B_\alpha) = MT^{-1/2} \sum_{t=1}^{T} S_t + o_p(1),$$

where $M = R[I, -Q_0 Q^{-1}]$ and $S_t = (q\psi_\theta(v_t), qv_t^* - u_t^*)' \otimes x_t$. Since x_t', u_t, v_t are α-mixing by assumption, and S_t is a measurable function of x_t', u_t, v_t, it follows that S_t is also α-mixing. Next, $E(S_t) = 0$ by Assumptions 3(iv) and 4(ii). Finally, Assumption 4(i) provides all the moment conditions necessary to invoke Theorem 5.20 of White (2001). Hence, we have:

$$V_T^{-1/2} T^{-1/2} \sum_{t=1}^{T} S_t \xrightarrow{d} N(0, I),$$

where $V_T = var\left(T^{-1/2}\sum_{t=1}^{T} S_t\right)$. Hence, the proof is completed. QED.

Proof of Lemma 1: With OLS first-stage estimators, we have $\sigma_0^2(q) = aq^2 + 2bq + c$, where $a = E\left[f(0)^{-1}\psi_\theta(v_t) - v_t^*\right]^2, b = E\left[\left(f(0)^{-1}\psi_\theta(v_t) - v_t^*\right)u_t^*\right]$ and $c = E(u_t^{*2})$, which corresponds to a convex parabolic curve that attains its minimum at

$$q^* = -\frac{b}{a} = \frac{E(v_t^* u_t^*) - f(0)^{-1}E(\psi_\theta(v_t)u_t^*)}{f(0)^{-2}\theta(1-\theta) + E(v_t^{*2}) - 2f(0)^{-1}E(\psi_\theta(v_t)v_t^*)},$$

which completes the proof. QED.

Proof of the TLS case: Since it is generated by censorship at two symmetric quantiles, the error term in the first step of the calculus of the TLS estimator can be written as $\tilde{v}_t = F_v^{-1}(\mu)I_{[v_t<F_v^{-1}(\mu)]} + F_v^{-1}(1-\mu)I_{[v_t<F_v^{-1}(1-\mu)]} + v_t I_{[F_v^{-1}(\mu)<v_t<F_v^{-1}(1-\mu)]}$. This transformation of the initial error term v_t corresponds to the trimming that is performed by using quantile regressions before applying the least square estimator.

The terms with a negative sign in Assumption $3''$ is what remains from the condition $E(\tilde{v}_t\,|x_{(j)t}, j = 2, ...) = 0$ that has not been cancelled, i.e. $E(\tilde{v}_t)$. For the OLS estimator, this is $(Ev_t, 0, ..., 0)$. For the TLS, this yields $(E(t\_v_t)+\mu(F_v^{-1}(\mu)+F_v^{-1}(1-\mu)), 0, ..., 0)$, where $t\_v_t = v_t I_{[F_v^{-1}(\mu)<v_t<F_v^{-1}(1-\mu)]}$ is the truncated error term. Indeed, using the above formula for $\tilde{v}_t$, we have

$E\tilde{v}_t = F_v^{-1}(\mu)P[v_t < F_v^{-1}(\mu)] + F_v^{-1}(1-\mu)P[v_t < F_v^{-1}(1-\mu)] + \int_{F_v^{-1}(\mu)}^{F_v^{-1}(1-\mu)} v f_v(v)dv$

$= F_v^{-1}(\mu)F[F_v^{-1}(\mu)] + F_v^{-1}(1-\mu)\{1 - F[v_t < F_v^{-1}(1-\mu)]\} + E(t\_v_t)$, which gives the result: $\left[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)\right]\mu + E(t\_v_t)$. QED.

Appendix B: Simulation Design

We base our simulations on a simultaneous equation system with two simple equations. The first equation, which is the equation of interest, contains two endogenous variables and two exogenous variables including a constant. Four exogenous variables are present in the whole system. The structural simultaneous equation system can be written $B \begin{bmatrix} y_t \\ Y_t \end{bmatrix} + \Gamma x_t = U_t$, 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 equal to one. The error term U_t is a 2×1 vector of error terms. We specify the structural parameters as follows: $B = \begin{bmatrix} 1 & -0.5 \\ -0.7 & 1 \end{bmatrix}$ and $\Gamma = \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & 0.2 \end{bmatrix}$. The system is over-identified by the exclusion restrictions $\Gamma_{13} = \Gamma_{14} = \Gamma_{22} = 0$. Moreover, $[\ v \quad V\ ] = U\,(B')^{-1}$. Hence, $\gamma = 0.5$ and $\beta' = (\beta_0, \beta_1) = (1, 0.2)$.

The choice of the parameter values is led by the following considerations. Only moderate cross effects of the two endogenous variables are specified so that the endogeneity problem be interesting but not extreme. Identification restrictions and the degree of over-identification drive the occurrence of exogenous variables in the equations. Moderate, while non-negligible and comparable effects are allowed for these variables.

The error v in the reduced-form equations is generated so as to satisfy Assumption 3(iv): $v = v^e - F_{v^e}^{-1}(\theta)$ where $v^e = \sigma(x_{5t})w_t$, w_t is generated by using alternatively the distributions N$(0,1)$, $t(3)$ and LN(0,1) with autocorrelation coefficient -0.1 and x_{5t} is generated from a distribution N$(0,1)$ independently of other random variables and errors. Because we assume that x_{5t} is independent of w_t and w_t is iid, $F_{v^e}^{-1}(\theta) = \sigma(x_{5t})F_w^{-1}(\theta)$, where $F_w^{-1}(\theta)$ is the inverse cumulative function of w_t evaluated at θ. The scale factor is $\sigma(x_{5t}) = 1 + \delta x_{5t}$. We choose $\delta = 0.05$ under heteroskedasticity and $\delta = 0$ under iid. The errors V_j are generated in the same way, albeit without heteroskedasticity. Then, we draw the second to fourth columns in X from the normal distribution with mean $(0.5, 1, -0.1)'$, variances normalized to 1, $cov(x_2, x_3) = 0.3, cov(x_2, x_4) = 0.1$ and $cov(x_3, x_4) = 0.2$, where x_2, x_3 and x_4 are respectively the second, third and fourth components of x_t. The correlations between the exogenous variables are neither extreme nor negligible. Given $X, [\ v \quad V\ ]$ and $[\ \pi_0 \quad \Pi_0\ ] = -\Gamma'(B')^{-1}$, we generate the endogenous variables $[\ y \quad Y\ ]$ by using the reduced-form equation: $[\ y \quad Y\ ] = X[\ \pi_0 \quad \Pi_0\ ] + [\ v \quad V\ ]$

Table 1(a). Simulation Means and Standard Deviations of $2SQR1(\theta, q=1)$: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-0.75	-0.35	-0.01	0.31	0.77
		Std	2.18	1.15	0.83	0.67	0.58
	$\tilde{\beta}_1$	Mean	0.01	0.00	0.00	0.00	0.00
		Std	0.35	0.23	0.21	0.23	0.35
	$\tilde{\gamma}$	Mean	-0.01	0.00	0.00	0.01	0.00
		Std	0.51	0.34	0.31	0.33	0.49
T = 300	$\tilde{\beta}_0$	Mean	-0.84	-0.34	-0.01	0.33	0.81
		Std	0.83	0.43	0.33	0.26	0.22
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.14	0.09	0.09	0.09	0.13
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.01
		Std	0.19	0.12	0.12	0.13	0.19

Table 1(b). Simulation Means and Standard Deviations of $2SQR1(\theta, q=q^*)$: N(0,1).

		θ (q^*)	0.05 (0.0013)	0.25 (-0.0003)	0.50 (0.0002)	0.75 (0.0003)	0.95 (0.0027)
T = 50	$\tilde{\beta}_0$	Mean	0.59	0.23	-0.01	-0.26	-0.62
		Std	1.19	0.89	0.71	0.54	0.36
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.19	0.19	0.18	0.18	0.19
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.27	0.26	0.26	0.26	0.27
T = 300	$\tilde{\beta}_0$	Mean	0.72	0.29	-0.01	-0.31	-0.74
		Std	0.44	0.34	0.27	0.21	0.14
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.07	0.07	0.07	0.07	0.07
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.10	0.10	0.10	0.10	0.10

Table 1(c). Simulation Means and Standard Deviations of $2SQR1(\theta, q=\hat{q})$: N(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	0.22	0.15	-0.01	-0.20	-0.26
		Std	1.49	0.91	0.72	0.54	0.40
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.22	0.19	0.18	0.19	0.22
	$\tilde{\gamma}$	Mean	0.01	0.01	0.00	0.01	0.01
		Std	0.33	0.26	0.26	0.27	0.32
	$\hat{q}$	Mean	0.19	-0.01	-0.05	0.07	0.31
		Std	0.33	0.23	0.20	0.20	0.20
T = 300	$\tilde{\beta}_0$	Mean	0.62	0.25	-0.01	-0.27	-0.62
		Std	0.46	0.34	0.27	0.21	0.16
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.07	0.07	0.07	0.07	0.07
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.10	0.10	0.10	0.10	0.10
	$\hat{q}$	Mean	0.08	0.00	-0.05	0.00	0.10
		Std	0.09	0.11	0.12	0.11	0.09

Table 2(a). Simulation Means and Standard Deviations of $2SQR1(\theta, q=1)$: t (3).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-1.07	-0.34	0.01	0.42	1.36
		Std	7.32	2.04	1.42	1.06	1.05
	$\tilde{\beta}_1$	Mean	0.04	0.02	0.01	-0.01	-0.01
		Std	0.88	0.33	0.28	0.34	0.85
	$\tilde{\gamma}$	Mean	-0.06	-0.02	-0.01	-0.01	-0.1
		Std	1.48	0.59	0.51	0.54	1.43
T = 300	$\tilde{\beta}_0$	Mean	-1.18	-0.40	-0.02	0.37	1.20
		Std	2.17	0.59	0.40	0.33	0.33
	$\tilde{\beta}_1$	Mean	0.02	0.00	0.00	0.00	0.00
		Std	0.29	0.11	0.10	0.12	0.30
	$\tilde{\gamma}$	Mean	0.00	0.00	0.01	0.01	0.01
		Std	0.43	0.17	0.14	0.17	0.42

Table 2(b). Simulation Means and Standard Deviations of $2SQR1(\theta, q=q^*)$: t (3).

		θ (q^*)	0.05 (-0.079)	0.25 (0.537)	0.50 (0.835)	0.75 (0.538)	0.95 (-0.078)
T = 50	$\tilde{\beta}_0$	Mean	0.90	0.01	0.02	0.10	-0.74
		Std	2.98	1.89	1.42	1.00	0.43
	$\tilde{\beta}_1$	Mean	0.02	0.02	0.01	0.01	0.02
		Std	0.35	0.32	0.28	0.31	0.34
	$\tilde{\gamma}$	Mean	-0.03	-0.02	-0.01	-0.02	-0.03
		Std	0.58	0.55	0.51	0.50	0.53
T = 300	$\tilde{\beta}_0$	Mean	0.87	-0.06	-0.02	0.02	-0.92
		Std	0.94	0.56	0.40	0.32	0.16
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.12	0.11	0.10	0.11	0.12
	$\tilde{\gamma}$	Mean	0.01	0.01	0.01	0.01	0.01
		Std	0.18	0.16	0.14	0.16	0.18

Table 2(c). Simulation Means and Standard Deviations of $2SQR1(\theta, q=\hat{q})$: t (3).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-0.19	0.10	0.07	-0.01	-0.22
		Std	13.6	2.02	1.66	1.04	0.86
	$\tilde{\beta}_1$	Mean	0.03	0.01	0.01	0.01	0.01
		Std	0.56	0.30	0.31	0.31	0.52
	$\tilde{\gamma}$	Mean	0.07	-0.01	-0.02	-0.02	-0.01
		Std	2.74	0.58	0.59	0.52	1.19
	$\hat{q}$	Mean	0.19	0.30	0.32	0.32	0.29
		Std	0.54	0.30	0.27	0.28	0.37
T = 300	$\tilde{\beta}_0$	Mean	0.80	0.01	-0.01	-0.04	-0.87
		Std	1.02	0.57	0.40	0.33	0.29
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.12	0.11	0.10	0.11	0.13
	$\tilde{\gamma}$	Mean	0.01	0.01	0.00	0.00	0.01
		Std	0.19	0.16	0.15	0.16	0.19
	$\hat{q}$	Mean	0.01	0.45	0.57	0.44	0.04
		Std	0.16	0.18	0.16	0.18	0.16

Table 3(a). Simulation Means and Standard Deviations of $2SQR1(\theta, q=1)$: LN(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-0.50	-0.35	-0.10	0.37	2.00
		Std	1.43	1.26	1.56	2.06	3.00
	$\tilde{\beta}_1$	Mean	0.02	0.02	0.02	0.01	0.03
		Std	0.21	0.21	0.30	0.48	1.73
	$\tilde{\gamma}$	Mean	-0.05	-0.05	-0.05	-0.05	-0.14
		Std	0.35	0.33	0.47	0.88	2.64
T = 300	$\tilde{\beta}_0$	Mean	-0.74	-0.57	-0.33	0.15	1.85
		Std	0.49	0.50	0.56	0.66	1.08
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.01	0.02
		Std	0.08	0.09	0.11	0.19	0.65
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.01	0.02
		Std	0.11	0.13	0.16	0.27	0.91

Table 3(b). Simulation Means and Standard Deviations of $2SQR1(\theta, q=q^*)$: LN(0,1).

		θ (q^*)	0.05 (1.0388)	0.25 (1.051)	0.50 (0.972)	0.75 (0.167)	0.95 (-0.146)
T = 50	$\tilde{\beta}_0$	Mean	-0.56	-0.41	-0.08	0.18	-1.25
		Std	1.41	1.25	1.56	1.57	0.91
	$\tilde{\beta}_1$	Mean	0.02	0.02	0.02	0.02	0.03
		Std	0.21	0.21	0.30	0.40	0.45
	$\tilde{\gamma}$	Mean	-0.05	-0.04	-0.05	-0.06	-0.07
		Std	0.35	0.33	0.47	0.65	0.71
T = 300	$\tilde{\beta}_0$	Mean	-0.79	-0.62	-0.31	-0.08	-1.42
		Std	0.49	0.50	0.56	0.54	0.32
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.08	0.09	0.11	0.16	0.18
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.01
		Std	0.11	0.13	0.16	0.22	0.24

Table 3(c). Simulation Means and Standard Deviations of $2SQR1(\theta, q=\hat{q})$: LN(0,1).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-0.39	0.04	0.18	0.21	-1.01
		Std	1.55	1.58	2.20	2.38	1.82
	$\tilde{\beta}_1$	Mean	0.02	0.02	0.02	0.02	0.02
		Std	0.22	0.25	0.34	0.40	0.75
	$\tilde{\gamma}$	Mean	-0.06	-0.05	-0.06	-0.07	-0.11
		Std	0.37	0.41	0.66	1.01	2.06
	$\hat{q}$	Mean	0.92	0.65	0.55	0.28	0.02
		Std	0.11	0.13	0.25	0.40	0.37
T = 300	$\tilde{\beta}_0$	Mean	-0.68	-0.40	-0.23	-0.03	-1.37
		Std	0.50	0.52	0.56	0.55	0.37
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.09	0.09	0.11	0.16	0.18
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.01
		Std	0.12	0.13	0.16	0.23	0.25
	$\hat{q}$	Mean	0.96	0.85	0.85	0.38	-0.17
		Std	0.04	0.04	0.06	0.31	0.08

Table 4(a). Simulated Standard Deviations of 2SQR1 $(\theta,\ \hat{q})$ and Cramer-Rao Bounds with T = 50

Estimator	First Stage	Second Stage		N(0,1)	t (3)	LN(0,1)
CR bounds			$\hat{\beta}_1$ $\hat{y}$	0.19 0.26	0.20 0.28	0.04 0.06
2SLS	LS	LS	$\hat{\beta}_1$ $\hat{y}$	0.21 0.30	0.41 0.78	0.42 0.70
2SQR $(\theta,\ \hat{q})$	LS	Quantile(θ =0.05)	$\hat{\beta}_1$ $\hat{y}$	0.22 0.33	0.56 2.74	0.22 0.37
2SQR $(\theta,\ \hat{q})$	LS	Quantile(θ =0.25)	$\hat{\beta}_1$ $\hat{y}$	0.19 0.26	0.30 0.58	0.25 0.41
2SQR $(\theta,\ \hat{q})$	LS	Quantile(θ =0.50)	$\hat{\beta}_1$ $\hat{y}$	0.19 0.26	0.31 0.59	0.34 0.66
2SQR $(\theta,\ \hat{q})$	LS	Quantile(θ =0.75)	$\hat{\beta}_1$ $\hat{y}$	0.19 0.27	0.31 0.52	0.40 1.01
2SQR $(\theta,\ \hat{q})$	LS	Quantile(θ =0.95)	$\hat{\beta}_1$ $\hat{y}$	0.22 0.32	0.52 1.19	0.75 2.06

Table 4(b). Simulated Standard Deviations of 2SQR2 (θ, q) with T = 50

Estimator	First Stage	Second Stage		N(0,1)	t (3)	LN(0,1)
Trim (φ) -2SQR $(\theta,1)$	Trim(0.25)	Quantile(θ =0.05)	$\hat{\beta}_1$ $\hat{y}$	0.36 0.52	0.82 1.22	0.15 0.22
Trim (φ) -2SQR $(\theta,1)$	Trim(0.25)	Quantile(θ =0.25)	$\hat{\beta}_1$ $\hat{y}$	0.24 0.37	0.31 0.44	0.17 0.25
Trim (φ) -2SQR $(\theta,1)$	Trim(0.25)	Quantile(θ =0.50)	$\hat{\beta}_1$ $\hat{y}$	0.23 0.36	0.24 0.39	0.23 0.35
Trim (φ) -2SQR $(\theta,1)$	Trim(0.25)	Quantile(θ =0.75)	$\hat{\beta}_1$ $\hat{y}$	0.24 0.37	0.31 0.48	0.42 0.63
Trim (φ) -2SQR $(\theta,1)$	Trim(0.25)	Quantile(θ =0.95)	$\hat{\beta}_1$ $\hat{y}$	0.35 0.51	0.79 1.14	1.63 2.33
Trim (φ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile(θ =0.05)	$\hat{\beta}_1$ $\hat{y}$	0.24 0.40	0.51 0.85	0.15 0.23
Trim (φ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile(θ =0.25)	$\hat{\beta}_1$ $\hat{y}$	0.21 0.36	0.26 0.39	0.18 0.27
Trim (φ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile(θ =0.50)	$\hat{\beta}_1$ $\hat{y}$	0.21 0.34	0.24 0.44	0.23 0.35
Trim (φ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile(θ =0.75)	$\hat{\beta}_1$ $\hat{y}$	0.22 0.36	0.26 0.40	0.29 0.46
Trim (φ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile(θ =0.95)	$\hat{\beta}_1$ $\hat{y}$	0.23 0.36	0.36 0.53	0.46 0.65

Table 5(a). Simulated Standard Deviations of 2SQR1$(\theta, \hat{q})$ and Cramer-Rao Bounds with T = 300

Estimator	First Stage	Second Stage		N(0,1)	t(3)	LN(0,1)
CR bounds			$\hat{\beta}_1$	0.07	0.08	0.02
			$\hat{\gamma}$	0.10	0.12	0.02
2SLS	LS	LS	$\hat{\beta}_1$	0.07	0.12	0.16
			$\hat{\gamma}$	0.10	0.18	0.22
2SQR$(\theta, \hat{q})$	LS	Quantile(θ=0.05)	$\hat{\beta}_1$	0.07	0.12	0.09
			$\hat{\gamma}$	0.10	0.19	0.12
2SQR$(\theta, \hat{q})$	LS	Quantile(θ=0.25)	$\hat{\beta}_1$	0.07	0.11	0.09
			$\hat{\gamma}$	0.10	0.16	0.13
2SQR$(\theta, \hat{q})$	LS	Quantile(θ=0.50)	$\hat{\beta}_1$	0.07	0.10	0.11
			$\hat{\gamma}$	0.10	0.15	0.16
2SQR$(\theta, \hat{q})$	LS	Quantile(θ=0.75)	$\hat{\beta}_1$	0.07	0.11	0.16
			$\hat{\gamma}$	0.10	0.16	0.23
2SQR$(\theta, \hat{q})$	LS	Quantile(θ=0.95)	$\hat{\beta}_1$	0.07	0.13	0.18
			$\hat{\gamma}$	0.10	0.19	0.25

Table 5(b). Simulated Standard Deviations of 2SQR2(θ, q) with T = 300

Estimator	First Stage	Second Stage		N(0,1)	t(3)	LN(0,1)
Trim(φ)-2SQR$(\theta,1)$	Trim(0.25)	Quantile(θ=0.05)	$\hat{\beta}_1$	0.14	0.28	0.05
			$\hat{\gamma}$	0.20	0.42	0.07
Trim(φ)-2SQR$(\theta,1)$	Trim(0.25)	Quantile(θ=0.25)	$\hat{\beta}_1$	0.09	0.11	0.06
			$\hat{\gamma}$	0.13	0.16	0.08
Trim(φ)-2SQR$(\theta,1)$	Trim(0.25)	Quantile(θ=0.50)	$\hat{\beta}_1$	0.09	0.09	0.09
			$\hat{\gamma}$	0.12	0.13	0.12
Trim(φ)-2SQR$(\theta,1)$	Trim(0.25)	Quantile(θ=0.75)	$\hat{\beta}_1$	0.09	0.11	0.17
			$\hat{\gamma}$	0.13	0.16	0.24
Trim(φ)-2SQR$(\theta,1)$	Trim(0.25)	Quantile(θ=0.95)	$\hat{\beta}_1$	0.13	0.30	0.62
			$\hat{\gamma}$	0.19	0.41	0.89
Trim(φ)-2SQR$(\theta, \hat{q})$	Trim(0.25)	Quantile(θ=0.05)	$\hat{\beta}_1$	0.08	0.10	0.05
			$\hat{\gamma}$	0.11	0.14	0.07
Trim(φ)-2SQR$(\theta, \hat{q})$	Trim(0.25)	Quantile(θ=0.25)	$\hat{\beta}_1$	0.08	0.09	0.06
			$\hat{\gamma}$	0.11	0.13	0.08
Trim(φ)-2SQR$(\theta, \hat{q})$	Trim(0.25)	Quantile(θ=0.50)	$\hat{\beta}_1$	0.08	0.09	0.09
			$\hat{\gamma}$	0.11	0.13	0.12
Trim(φ)-2SQR$(\theta, \hat{q})$	Trim(0.25)	Quantile(θ=0.75)	$\hat{\beta}_1$	0.08	0.09	0.12
			$\hat{\gamma}$	0.11	0.13	0.17
Trim(φ)-2SQR$(\theta, \hat{q})$	Trim(0.25)	Quantile(θ=0.95)	$\hat{\beta}_1$	0.08	0.10	0.10
			$\hat{\gamma}$	0.11	0.14	0.14