Bias Transmission and Variance Reduction in Two-Stage Estimation

Tae-Hwan Kim
School of Economics, University of Nottingham
University Park, Nottingham NG7 2RD, UK
(Phone) 44-115-951-5466 (Fax) 44-115-951-4159
Tae-Hwan.Kim@nottingham.ac.uk

Christophe Muller
Departamento de Fundamentos del Analisis Economico, Universidad de Alicante,
Campus de San Vicente, 03080, Alicante, Spain
(Phone) 34-965-90-3614 (Fax) 34-965-90-3898
cmuller@merlin.fae.ua.es

November 2003

Abstract: In this paper, we study the transmission of an asymptotic bias in two-stage regressions with non-iid errors and random regressors in which the possible endogeneity of some explanatory variables is treated via first-stage predictive equations. In particular, we characterise the transmission of an asymptotic bias in the first-stage estimates to the second-stage estimates. As an example, we fully develop the case of two-stage quantile regressions. Finally, Monte Carlo simulation results illustrate the occurrence of the bias in small samples.

JEL Codes: C13, C30.

Key Words: Two-Stage Estimation, Quantile Regression, Endogeneity, Asymptotic Bias, Semi-Parametric Estimators.

We are grateful to R. Davidson, W. Newey, J. Powell, H. White, O. Linton and P.C.B. Phillips for useful discussions about a preliminary version of this paper and to participants in presentations in the University of Nottingham, CREST-INSEE in Paris, London School of Economics, University of California at San Diego, the 1999 Econometric Study Group Meeting in Bristol, the 2000 Royal Statistical Society Conference in Reading and the 2002 Afriecon Conference in Kruger Park for their comment. Remaining errors are ours. We acknowledge the Award No. 30,649 from the British Academy.

1 Introduction

The equations that researchers want to estimate often describe relations in which some independent variables are endogenous. In linear models estimated with least squares methods, the usual response to such a situation is to replace the endogenous regressors with predictions using ancillary equations based on other exogenous variables and to modify the formula of the asymptotic covariance matrix of estimates. The most famous method is the two-stage least squares estimator of which conditions for consistency and asymptotic normality are known.[1] 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.[2]

There are many situations where the ancillary first-stage regressions, used for predicting the endogenous regressors, may be asymptotically biased: missing variables, invalid instrumental variables, model misspecification, faulty estimation methods, etc. Many of these estimation methods rely on semi-parametric restrictions for the first-stage equations. In general nothing guarantees that such restrictions are rigourously satisfied. Often when there is an intercept term in the model, one may hope that if the imposed semi-parametric restrictions are not satisfied (e.g., that the error term has non-zero mean), this may only affect the consistency of the intercept estimator in the first-stage equation.

However, it is generally the second stage of the estimation that includes the results of interest for applied researchers. When the first stage estimation method is asymptotically biased, how the bias is transmitted to the second stage of the estimation is unclear in realistic settings with random and endogenous regressors and non iid errors. In this paper, we investigate the channel of this bias transmission by eliciting the algebraic structure of the asymptotic representation of the two-stage estimator. Notably, we show that an asymptotic bias on the intercept of the first-stage estimation is exclusively and integrally conveyed to the intercept in the second stage of the estimation.

This state of affairs has several interesting consequences. First, the choice of the first-stage and second-stage estimation methods matters if one worries about generation of asymptotic biases, even if only on the intercept coefficient. We will show that there are cases where the semi-parametric restrictions imposed on the two stages may lead to contradictions that result in the occurrence of an asymptotic bias for the two-stage estimator.

Second, we shall show that the estimator of the intercept coefficient in the

[1]Malinvaud (1970), Amemiya (1985).

[2]Heckman (1976), Newey (1985), Pagan (1986), Newey (1989), Newey (1994).

second stage may be inconsistent, but the estimators of the slope coefficients will be consistent in plausible cases, which may be all that interest the applied researcher. However, even in that case there are still precautions to take for inference. Indeed, the inconsistent intercept estimator is present in the residual. Then, usual 'plug-in' methods employed for estimating the covariance matrix of the parameters by replacing error terms in the formula of this matrix by residuals may yield inconsistent estimators. One solution to this issue is to correct the estimator of this covariance matrix to account for the bias on the intercept. This is generally easy to do, as soon as the applied researcher is aware of the problem.

Finally, we emphasize that the result of the integral transmission of the bias on the first-stage intercept to the second-stage intercept is valid in very general settings, with random endogenous regressors and non independent and non identically distributed errors. To the best of our knowledge, such property has only been noticed, except under very restrictive setting. In particular, we shall show that the traditional approach of analysing the bias on the intercept for quantile regressions does not provide any insight for general settings.

Our main example is that of two-stage quantile regressions of the linear model. Focusing on quantile regressions has several advantages. First, it enables us to limit our attention to a popular estimation method, rather than losing ourselves into vague generalities. Second, we provide the complete inference package for two-stage quantile regression, which was not available before.

Quantile regression and least absolute deviations estimators have recently become very popular estimation methods. Quantile regressions have been used for studying wages and living standards[3], firm data[4], financial data[5], and longitudinal data.[6] They are often chosen for two kinds of properties. Firstly, they provide robust estimates, particularly for misspecification errors related to non-normality, but also for the presence of outliers. Secondly, they allow the researcher to concentrate her attention on specific parts of the conditional distribution of the dependent variable.

The theoretical literature on quantile regression and LAD estimators is extensive since the seminal paper by Koenker and Bassett (1978). The asymptotic behaviour of these estimators has been extensively studied.[7] A few extensions exist for two-stage estimators. Amemiya (1982) and Powell (1983) have treated the case of the two-stage least absolute deviations (2SLAD). Chen and Portnoy (1996) study two-stage quantile regressions

[3]Buchinsky (1995, 98), Jalan and Ravallion (1998), Machado and Mata (2001).

[4]Mata and Machado (1996).

[5]Engle and Manganelli (1999), Granger and Sin (2000).

[6]Lipsitz et al. (1997).

[7]Koenker and Bassett (1978, 82), Bassett and Koenker (1978, 86), Powell (1983), Weiss (1990), Phillips (1991), Pollard (1991).

where the first-stage estimators are trimmed least squares estimators and LAD estimators, although only under assumption of symmetric iid error distributions. However, up to now, no general study of two-stage quantile regression estimators is available for general first-stage estimators, and general assumptions on error terms and regressors.

Although it does not correspond to our simple two-stage approach, the possibility of dealing with endogeneity problems in quantile regressions has already been examined in the literature. Least-absolute-error-difference estimators for a single equation from a simultaneous equation model have been studied by Kemp (1999) and Sakata (2001). A quantile treatment effects estimator has been proposed by Abadie, Angrist and Imbens (2002) by solving a convex programming problem with a preliminary non-parametric estimation of a nuisance function. MaCurdy and Timmins (2000) use an estimator for ARMA models and quantile regressions. Kim and Muller (2003) deal with the case where the same quantile regression method is employed for the two stages, in the iid case, and thereby avoid the bias transmission issue.

In this paper, we study the asymptotic and small sample properties of general two-stage linear regression estimators. Moreover, we offer specific contributions for two-stage quantile regressions. Firstly, we generalise the 2SLAD results to estimators that enables one to focus on different parts of the conditional distribution of the dependent variable. Secondly, we provide results with random exogenous variables and dependent and non identically distributed error terms. Thirdly, we clarify the link between the assumptions for first-stage and second-stage errors. In particular, we deal with biases in the first stage of this estimation. This is all the more important that the case without bias corresponds to restrictive conditions on error terms when considering arbitrary quantiles. Usual renormalisations of the intercept are generally not sufficient to eliminate the bias. Fourthly, we conduct Monte Carlo simulations that provide insight on small sample properties and illustrate the role of different parameters of the problem.

Section 2 discusses the model and the assumptions. In Section 3, we derive the asymptotic representation of the two-stage quantile regression estimators. We discuss the asymptotic bias for general two-stage estimators in Section 4. We analyse in Section 5 the asymptotic normality and the asymptotic covariance matrix of two-stage quantile regression based on LS predictions. We present simulation results in Section 6. Finally, Section 7 concludes. All proofs are in Appendix A.

2 The Model

We are interested in the parameter (α_0) in an equation that is given in the following matrix form for a sample of T observations:

$$\begin{aligned} y &= X_1\beta_0 + Y\gamma_0 + u \\ &= Z\alpha_0 + u \end{aligned} \tag{1}$$

where $[y, Y]$ is a $T \times (G+1)$ matrix of endogenous variables, X_1 is $T \times K_1$ matrix of exogenous variables, $Z = [X_1, Y]$, $\alpha_0' = [\beta_0', \gamma_0']$, and u is a $T \times 1$ vector. We denote by X_2 the matrix of $K_2(= K - K_1)$ exogenous variables absent from (1). Let us assume that Y can be predicted from the set of exogenous variables:

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

where $X = [X_1, X_2]$ is a $T \times K$ matrix, Π_0 is a $K \times G$ matrix of unknown parameters and V is a $T \times G$ matrix of unknown error terms. We assume that the first column of X_1 is a vector of 1's. Using (1) and (2), y can also be expressed from the exogenous variables:

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

where $\pi_0 = \left[\begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi_0\right]\alpha_0 = H(\Pi_0)\alpha_0$ and by definition $v = u + V\gamma_0$. Equations (2) and (3) are the basis of the first-stage estimation that yields some estimators $\hat{\pi}$, $\hat{\Pi}$ respectively of π_0, Π_0. We now specify the data generating process.

Assumption 1 *The sequence $\{(x_t', u_t, v_t)\}$ is strong mixing with mixing numbers $\{\alpha(s)\}$ of size $-(2K+1)(K+1)$. Here, x_t, u_t and v_t are the t^{th} elements in X, u and v respectively.*[8]

Studying quantile regressions with dependent processes is unusual, although some interesting results are in Weiss (1990). As we mentioned before, the asymptotic properties of the two-stage estimator can be of interest when the first-stage estimators $\hat{\pi}$ and $\hat{\Pi}$ are asymptotically biased. This situation arises for example when the LS estimation method is used in the first stage

[8]The sequence $\{W_t\}$ of random variables is strong 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 where 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 strong mixing of size $-a$ if $\alpha(s) = O(s^{-a-\varepsilon})$ for some $\varepsilon > 0$.

and the quantile regression method is used in the second stage. Considering asymptotically biased first-stage estimators is interesting on two grounds. First, the first-stage estimates are sensitive to various misspecifications and estimation difficulties that may generate asymptotic biases. Second, when no intercept term is present in the first-stage equation, the usual normalisation approach is not possible and asymptotic biases may occur as we shall show later on. To be able to account for inconsistent first-stage estimation, we use the following assumption.

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* $T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi) = O_p(1)$*, where* $\|a\| = (a'a)^{1/2}$.

The bias terms B_π and B_Π may not be bounded under fat-tailed error distributions if OLS is used in the first stage. Therefore, imposing Assumption 2 excludes some error distributions.

We now turn to the optimisation program from which the two-stage estimator is calculated. To save on space, we explicitly develop only the case of the Two-Stage Quantile Regression, while the generalisation to general two-stage M-estimators is obvious. Following the literature on quantile regressions, we define $\rho_\theta : R \to R^+$ for a given $\theta \in (0,1)$ as $\rho_\theta(z) = z\psi_\theta(z)$, where $\psi_\theta(z) = \theta - 1_{[z\leq 0]}$ and $1_{[.]}$ is the Kronecker index.

The motivation for the two stage approach is to deal with an endogeneity problem. If the first-order conditions, $\sum_{t=1}^{T} Z_t\psi_\theta(y_t - Z_t'\alpha) = o_p(1)$, were satisfied, then the one-stage quantile regression estimator would be consistent. However, when u and Y are correlated, which occurs under the non-separability in parameters of the joint density due to the endogeneity of Y, these first-order conditions are not satisfied. Therefore, the first-stage quantile regression estimator of α_0 is generally not consistent.

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 programme.

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

where y_t is the t^{th} elements in y and q is a non-zero constant. The reformulation of the dependent variable as $qy_t + (1-q)x_t'\hat{\pi}$ has been introduced by Amemiya to improve efficiency by choosing parameter q.

In the next section, we discuss the asymptotic representation of the 2SQR(θ, q). We shall show that the following conditions are sufficient for the asymptotic representation.

Assumption 3 (i) $H(\Pi_0)$ *is of full column rank.*

(ii) *Let* $f_t(\lambda|x) = \frac{\partial}{\partial\lambda}F_t(\lambda|x)$ *be the conditional pdf and* $F_t(\lambda|x)$ *be the conditional cdf of* v_t*. It is assumed that* $f_t(\cdot|x)$ *is Lipschitz continuous for all* x*, strictly positive, and bounded; that is, there exists a constant* f_0 *such that* $0 < f_t(\cdot|x) < f_0$ *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) $E(\psi_\theta(v_t)|x_t) = 0$.
(v) $\sup_{t\geq 1} E(\|x_t\|^3) < C < \infty$ *for some positive constant* C*.*

Assumptions 3(i)-(iii) are standard in the literature. Note, however, that the conditional pdf $f_t(\cdot|x)$ may change with observation t. Assumption 3(iv) is the assumption that zero is the θ^{th}-quantile of the conditional distribution of v_t.[9] When there is an intercept term in the model, Assumption 3(iv) can be considered as an identification condition on the coefficient of the intercept. Indeed, $E(\psi_\theta(v_t)|x_t) = 0$ and $E(\psi_\theta(v_t)|x_t) \neq 0$ correspond to isomorphic statistical structures that distinguish themselves only by the value of the intercept term. They are observationally equivalent structures. Therefore, it is possible to impose $E(\psi_\theta(v)|x_t) = 0$, and thus to fix the value of the intercept, without loss of generality. Jureckova (1984) mentions that the non-existence of an intercept would affect the large sample properties of quantile regressions. This suggests that having a close look at this intercept is interesting. We shall show that the use of the intercept term for normalisation does not extend to the two-stage estimators because contradictions may occur between the semi-parametric restrictions in the first and the second stages. Assumption 3(v) is necessary for obtaining the stochastic equicontinuity of our empirical process of interest in the strong mixing case. We are now ready to study the asymptotic properties of the 2SQR(θ, q).

3 The Asymptotic Representation

The first step of the analysis is the derivation of an asymptotic representation of the 2SQR(θ, q). For this, we define an empirical process given by

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

where Δ is a $K\times 1$ vector. A direct application of Theorem II.8 in Andrews (1990) yields the following lemma. The lemma is proven only for the quantile

[9] Note that 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 1(iv), $F^{-1}(\theta)$ is now zero so that we have $f(0)^{-1}$ instead in the iid case.

regression case, but similar derivations can be done for general two-stage M-estimators.

Lemma 1 *Suppose that Assumptions 1 and 3 hold. Then, we have for any* $L > 0$,

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

We combine Lemma 1 and Assumption 2 to obtain the following asymptotic representation for the 2SQR(θ, q) with a possible bias.

Proposition 1 *Suppose that Assumptions 1-3 hold. Then, the 2SQR(θ, q) has the asymptotic representation*

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

where $B_\alpha = RQ_0\{(1-q)B_\pi - B_\Pi \gamma_0\}$, $R = Q_{zz}^{*-1} H(\Pi_0^*)'$, $Q_{zz}^* = H(\Pi_0^*)' Q_0 H(\Pi_0^*)$ *and* $\Pi_0^* = \Pi_0 + B_\Pi$.

This asymptotic representation shows that the asymptotic distribution of the second-stage estimator $T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$ depends on the asymptotic distribution of the first-stage estimators $T^{1/2}(\hat{\pi} - \pi_0 - B_\pi)$ and $T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0$.[10] Naturally, if $q = 1$, the influence of $\hat{\pi}$ disappears. The asymptotic representation of the 2SQR(θ, q) is composed of three additive terms. The first term in the right-hand-side 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 in the right-hand-side term correspond to the respective contributions of $\hat{\pi}$ and $\hat{\Pi}$ to the uncertainty of the estimator. Because of these contributions, contradictions between semi-parametric restrictions used in the first stage and the second stage may occur and yield biases that cannot be eliminated by renormalisation of the intercepts. We now discuss the issue of the asymptotic bias in the quantile regression case.

4 Asymptotic Bias

As for most estimation methods, an incorrect specification of the first stage in (2) and (3) may degrade the properties of the 2SQR(θ, q). However,

[10] Other derivations of asymptotic representations of quantile regression estimators have been developed (Phillips, 1991, Pollard, 1991), which involve slightly different assumptions. They have not been applied to 2SQR estimators.

some misspecifications of the first stage do not affect the estimation results of interest. In particular, we now show that an asymptotic bias on the intercept of the first stage estimator can be appropriately dealt with.

This result has not been exhibited in the literature. One reason for such lacuna may be that in the traditional approach of examining the conditional quantile defined as the inverse of the conditional distribution function, it is not obvious how the bias is transmitted to the two-stage estimator. In contrast, our analysis is based on the algebraic structure of the asymptotic representation of the two-stage estimator. This representation implicitly includes a projection that conveys the asymptotic properties of the first-stage estimators to the two-stage estimator. We shall show that this implies that asymptotic biases on the intercepts of the first-stage estimators affect only the intercept of the two-stage estimator.

This situation has several interesting implications. First, one should be careful when choosing the first-stage and second-stage methods in this type of two stage estimations. Without co-ordinating the semi-parametric restrictions at the two stages, asymptotic biases may occur as in the case of two-stage quantile regressions that we develop. Second, the sufficient stochastic assumptions to obtain the transmission of the bias on the first-stage intercept coefficients to the second-stage intercept only are very general, including the possibility of general serial correlations and homoscedasticity, and of endogenous variables in the equation of interest. Third, the possible presence of an asymptotic bias on the intercept of the two-stage estimator is more serious than it may first appears. Indeed, because residuals are often used for estimates of covariance matrices and of test statistics, an asymptotic bias on the intercept coefficient may lead to inconsistent inferences. In that sense, what is at stake here is not only the interpretation of one coefficient of the model, but also the danger of doing incorrect inferences based on the whole model.

We now turn to the explicit analysis of the asymptotic bias. Because most interesting results only arise if the asymptotic bias of the first-stage estimators exclusively affect the intercept term, we focus on the case where the first-stage estimators of the slope coefficients are consistent. According to the asymptotic representation in Proposition 1, biases in $\hat{\pi}$ and in $\hat{\Pi}$ are transmitted to the 2SQR(θ, q) through the matrix RQ. Just looking at the matrix RQ does not make obvious that the asymptotic bias in the first-stage estimators only affects the intercept coefficient of the second-stage estimator. This feature occurs for many two-stage estimation procedures that share the same algebraic structure for the asymptotic representation.

To isolate the intercept of the first-stage estimators, we decompose both matrix Q_0 and the first-stage estimator: $Q_0 = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ where Q_1 is the first column of Q_0 and Q_2 is a $K \times (K-1)$ matrix consisting of the remaining columns of Q_0, and $\hat{\pi} - \pi_0 = \begin{bmatrix} \hat{\pi}_{(1)} - \pi_{0(1)} \\ \hat{\pi}_{(2)} - \pi_{0(2)} \end{bmatrix}$, where $\hat{\pi}_{(1)}$

is the estimator of the constant coefficient. This yields $RQ_0(\hat{\pi} - \pi_0) = RQ_1(\hat{\pi}_{(1)} - \pi_{0(1)}) + RQ_2(\hat{\pi}_{(2)} - \pi_{0(2)})$, where the second term in the right-hand-side term is asymptotically unbiased by assumption. The contribution of the first-stage estimate $\hat{\Pi}$ can be similarly decomposed, and we do not repeat the calculus that is similar to the one for $\hat{\pi}$. It is therefore necessary and sufficient to study the product RQ_1 to understand the generation of a possible asymptotic bias of $\hat{\alpha}$. The next proposition presents our main result.

Proposition 2 *Given that* $R = [H(\Pi_0^*)'Q_0H(\Pi_0^*)]^{-1}H(\Pi_0^*)'$ *and* $H(\Pi_0^*) = \left[\begin{pmatrix} I_{K_1} \\ 0_{K_2 \times K_1} \end{pmatrix}, \; \Pi_0^* \right]$, *we have*

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

Proposition 2 implies that the only coordinate of $\hat{\alpha}$ for which there is a possible asymptotic bias in that case 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 the intercept in $\hat{\Pi}\gamma_0$.

Several favourable situations may occur. First, empirical researchers are generally interested in the slope components of $\hat{\alpha}$ rather than in its intercept coefficient. Then, any first-stage 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 $T^{1/2}(\hat{\Pi} - \Pi_0)$ 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 $q = 1$ guarantees that this bias disappears. Naturally, the first-stage estimation method can also be chosen to eliminate the biases on $\hat{\pi}$ and $\hat{\Pi}$ (e.g., by using the same quantile regressions in the two-stages as in Kim and Muller, 2003). However, we consider in this paper that the researcher chooses the first-stage estimation method freely for her own reasons, for example because there exists already some available estimation results.

We need to return to the normalisation of the model to assess the consequence of the choice of the estimation procedure. Assumption 3(iv) normalises the intercept on the θth quantile of the distribution of v_t. However, the intercept may be interesting in more than one quantile. In that case, two natural approaches are possible. First, the researcher may choose to use different adapted first stage methods for different quantiles, for example different quantile regressions with the same quantile as in the second stage. Second, she may alternatively decide to stick to the same first-stage estimation results for all the different second stage quantiles. For example, a least squares estimator or a least absolute deviations estimator may be systematically used for the first stage. This has the advantage of requiring only one trial of first-stage estimation, but it implies that the researcher is ready to accept the occurrence of the asymptotic bias on the intercept for almost all quantiles.

The traditional approach in the quantile regression literature[11] of directly deriving the intercept term from the normalisation assumption is convenient only in the case where the error terms are independent of the independent variable. Suppose we have $y_t = x_t'\beta + \varepsilon_t$ where the first element of x_t is one and ε_t is independent of x_t. Then, in this case the θ^{th} conditional quantile of y_t is $q_\theta(y_t|x_t) = x_t'\beta + F^{-1}(\theta)$, where F is the cdf of ε_t. Indeed, by definition $q_\theta(\varepsilon_t|x_t) = F^{-1}(\theta)$ and the variables x_t, do not perturb[12] the arrangement of the different quantiles of y. In contrast, when the regressors and the errors are not independent, the role of the normalisation is much more important than it appears at first sight. Indeed, in general it is not obvious how to calculate the translated intercept. This is because the translation depends on the joint distribution of error terms and exogenous variables, and this distribution may be characterised by heteroscedasticity and serial correlations. Furthermore, it is not obvious in the general case that the asymptotic bias is only on the intercept.

More explicitly, for our main model of interest, let $q_\theta(y_t \mid Y_t, x_{1t})$ be the θ^{th} conditional quantile of y_t given Y_t and x_{1t} where Y_t' and x_{1t}' are the t^{th} elements of Y and X_1 respectively. Then, we have

$$\int_{-\infty}^{q_\theta(y_t|Y_t,x_{1t})} f_{y_t|Y_t,x_{1t}}(y \mid Y_t, x_{1t})\, dy = \theta,$$

which implies that $q_\theta(y_t|Y_t, x_{1t}) = x_{1t}'\beta_0 + Y_t'\gamma_0 + F^{-1}_{u_t|Y_t,x_{1t}}(\theta)$. On the other hand, with first-stage estimators $\hat{Y}_t$ we have $q_\theta(y_t|\hat{Y}_t, x_{1t}) = x_{1t}'\beta_0 + \hat{Y}_t'\gamma_0 + F^{-1}_{u_t|\hat{Y}_t,x_{1t}}(\theta)$. If the first-stage estimators are OLS, then $\hat{Y}_t = x_t(X'X)^{-1}X'Y$ and $q_\theta(y_t|\hat{Y}_t, x_t) = x_{1t}'\beta_0 + x_t(X'X)^{-1}X'Y\gamma_0 + F^{-1}_{u_t|x_t(X'X)^{-1}X'Y,x_{1t}}(\theta)$. None of these expressions seems to provide much insight about the nature of the possible asymptotic biases because of the presence of $F^{-1}_{u_t|x_t(X'X)^{-1}X'Y_t,x_{1t}}(\theta)$. Whether this term can affect only the intercept term depends on the random association of the u_t, x_{1t} and $\hat{Y}_t$. This reasoning can be extended to other two-stage estimation methods by using the appropriate inversion procedure instead of the inverse cdf that is specific to the calculus with regression quantiles.

Therefore, it seems at first sight that nothing guaranteed a priori that an asymptotic bias would not generally appear on slope coefficients. We have

[11] A common approach in the literature is to normalise the model on a measure of central tendency (Koenker and Bassett, 1978).

[12] By definition of the independence, the parameters of the marginal distribution of the x_t can be factorized in the joint distribution, and disappear in the conditionning.

examined this issue by deriving the asymptotic representation of the estimator of interest, then by exhibiting a matrix identity appearing in this representation, and finally by exploiting this identity to show how an asymptotic bias on the intercept of the first-stage estimators is integrally transmitted to the intercept of the two-stage estimator. We now provide a direct intuition of the result.

The problem can be seen as understanding the term $F_{u_t|x_t(X'X)^{-1}X'Y_t,x_{1t}}$ when x_t is exogenous. If x_t is strictly exogenous, i.e. if there is separation of the marginal distribution of the x_t in the joint distribution, then it can be shown directly that $x_t(X'X)_{-1}X'Y_t$ is also strictly exogenous. This is because $X(X'X)^{-1}X'Y$ is the projection of Y_t the space spanned by X, and is therefore a fixed linear combination of strictly exogenous variables. Moreover, the term corresponding to observation t can be written as $x_t(X'X)^{-1}X'Y_t$ since only the t^{th} line of $X(X'X)^{-1}X'$ matters for calculation the projection corresponding to this observation. Then, in this case $F_{u_t|x_t(X'X)^{-1}X'Y_t,x_{1t}} = F_{u_t}$.

Now, if we relax the assumption of strict exogeneity in for example orthogonality with the error term, we have $u_t \perp x_t(X'X)^{-1}X'Y_t$ and x_{1t} with the same reasoning as above. Clearly, in that case $E[u_t|x_t(X'X)^{-1}X'Y_t, x_{1t}] = 0$ but it is not necessary that $q_\theta[u_t|x_t(X'X)^{-1}X'Y_t, x_{1t}] = 0$. However, under these assumptions we do not have $q_\theta[u_t|x_t] = 0$. Moreover, this orthogonality for OLS is not the type of exogeneity that we defined at the beginning. If instead we start from $q_\theta[u_t|x_t] = 0$, the above projection will ensure that $q_\theta[u_t|x_t(X'X)^{-1}X'Y_t, x_{1t}] = 0$ since the conditioning is nothing else than a special case of x_t and our definition of exogeneity is still based on an orthogonality condition. We now turn to the explicit derivation of the asymptotic covariance matrix for 2SQR(θ, q) with LS predictions.

5 Asymptotic Normality and Covariance Matrix with LS Predictions

In this section, we investigate the use of LS estimation for π_0 and Π_0 in the first step of 2SQR(θ, q). Naturally, if one is interested in robustness, it is a bad idea to use LS estimators in the first stage. However, there are several reasons to consider this case. First, the researcher may want to use quantile regressions not for their robustness but rather for the possibility of focusing on different locations of the conditional distribution of the dependent variable. Also, LS estimators are popular and available LS estimation results for the first-stage equations could be ready to be used. Moreover, one may be interested in robustness issues arising only from the second-stage setting, e.g., outliers for u. Then, using LS estimators as a first stage may improve the efficiency of the estimation procedure. Finally, that is what some empirical researchers do and are willing to do in practice. Some empirical studies[13] adopt first-stage least squares estimators with the second stage based on quantile regression. Therefore, it can be important for applied researchers to know the theoretical consequences of that approach, and we provide an answer in this section.

Using the LS estimation in the first stage yields consistency only for the slope coefficients of the 2SQR(θ, q). This is because the necessary condition $E(v_t) = 0$ for consistently estimating the intercept coefficient in (3) by OLS is not compatible with Assumption 3(iv). This problem has hardly been noticed in the literature, although this might be the reason why authors imposed symmetry of error terms (as in Chen, 1988, and in Chen and Portnoy, 1996). If $\theta = 1/2$ and the distribution is symmetric, then the bias vanishes, as in Powell (1983).

First, we define $V_t^* = V_t - E(V_t)$ and $v_t^* = v_t - E(v_t)$. Then, the reduced forms for Y_t and y_t in (2) and (3) can be expressed as

$$Y_t = x_t'\Pi_0^* + V_t^* \tag{5}$$

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

$$y_t = x_t'\pi_0^* + v_t^* \tag{6}$$

where $\pi_0^* = \pi_0 + B_\pi$ and $B_\pi = [E(v_t), 0, ..., 0]'_{(K\times 1)}$. We also define $u_t^* = v_t^* - V_t^*\gamma_0$, where it can be shown that $u_t^* = u_t - E(u_t)$. By construction, we have $E(V_t^*) = E(v_t^*) = E(u_t^*) = 0$.

Let $\tilde{\Pi}$ and $\tilde{\pi}$ be the LS estimators based on (5) and (6) respectively; that is we have

$$\begin{aligned} T^{1/2}(\tilde{\pi} - \pi_0^*) &= Q^{-1}T^{-1/2}\sum_{t=1}^{T} x_t v_t^* + o_p(1) \\ T^{1/2}(\tilde{\Pi} - \Pi_0^*) &= Q^{-1}T^{-1/2}\sum_{t=1}^{T} x_t V_t^* + o_p(1). \end{aligned}$$

Let $\tilde{\alpha}$ be the 2SQR(θ, q) based on the LS estimators $\tilde{\Pi}$ and $\tilde{\pi}$ in the first stage. By plugging the above expressions into the formula in Proposition 1, we obtain the asymptotic representation for the 2SQR(θ, q) based on LS predictions as follows;

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

[13] Arias et al. (2001), Garcia et al. (2001).

where $B_\alpha = RQ_0\{(1-q)B_\pi - B_\Pi\gamma\}$. Owing to Proposition 2 and the definitions of B_π and B_Π, we have $B_\alpha = ((1-q)E(v_t) - E(V_t)\gamma_0, 0, \ldots, 0)'$. The formula of the bias term B_α shows that the intercept estimator may be asymptotically biased, while the slope estimators are not, for usual semiparametric assumptions affecting only the location of the error distributions. Even when there is a bias, the asymptotic normality of $\tilde{\alpha} - \alpha_0 - B_\alpha$ can be easily derived. For this purpose, we impose the following assumptions.

Assumption 4 (i) *The sequence* $\{(u_t, v_t)\}$ *satisfies the following moment conditions that there exist finite constants* Δ_u *and* Δ_v *such that* $E|x_{ti}u_t^*|^3 < \Delta_u$ *and* $E|x_{ti}v_t^*|^3 < \Delta_v$ *for all* i *and* t.
(ii) $E(u_t^*|x_t) = 0$ *and* $E(v_t^*|x_t) = 0$.
(iii) *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* $S_t = (q\psi_\theta(v_t), qv_t^* - u_t^*)' \otimes x_t$ *and* $\otimes$ *is the Kronecker product.*

Proposition 3 *Suppose that Assumptions 1,3-4 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 normality of the slope coefficients is easily derived using Proposition 3 by truncating the vector of parameters. Let $\alpha_{0(1)}$ and $\alpha_{0(2)}$ be the intercept and slope coefficients respectively. We also decompose the 2SQR(θ, q) accordingly: $\tilde{\alpha}' = (\tilde{\alpha}_{0(1)}, \tilde{\alpha}'_{0(2)})$. Under the same conditions as in Proposition 3, we have $N_T^{-1/2}T^{1/2}(\tilde{\alpha}_{0(2)} - \alpha_{0(2)}) \xrightarrow{d} N(0, I)$, where $N_T = M_2V_TM_2'$ and M_2 is the last $(K_1 + G - 1)$ rows in M.

At this stage, we have shown that it is possible to obtain useful asymptotic properties of the 2SQR(θ, q) for a given value of q. Now, the main interest of introducing parameter q in the problem is to provide an opportunity to improve efficiency. This can be done trying several values of q or using prior information of values of q that worked well for past estimations. One may also want to adopt a more systematic approach and replace q by its optimal value obtained by minimising the asymptotic covariance matrix for which we derived an explicit expression. We now discuss such estimator of q and the impact that it may have on the 2SQR(θ, q).

The estimation of parameter q as a minimand of the asymptotic covariance matrix of the 2SQR(θ, q) raises a series of difficulties. First, there is no unique way of minimising a covariance matrix when the dimension is greater then one. One possibility is to select a matrix norm (e.g., trace or determinant) that will be minimised. Another one is to focus on one coefficient of interest, for example the coefficient of the return to education in wage equations, and to minimise only the asymptotic standard error for this coefficient. Moreover, in the iid case, unique values of q^* can be reached. Second, in general, no explicit formula may be available for q^*. This would make the whole estimation process less straightforward and implies to use numerical estimation techniques.

In the case of least-squares plus quantile regression estimation, using $q = 1$ and $E(V_t) = 0$ in the asymptotic representation in Proposition 1 would allow the researcher to avoid the occurrence of an asymptotic bias. By contrast, using different values for q would introduce an asymptotic bias isolated in the intercept term; but the asymptotic variance of the consistent slope estimates can be reduced. Since the bias can be easily corrected, one should try to improve efficiency in two-stage estimations whose results are often insufficiently accurate for the needs of applied researchers.

We focus on a case where an explicit formula for an estimator fo q, $\hat{q}$, can be exhibited, which allows us to convey the intuition of the DGP features driving the estimation properties. Finally, we shall present Monte Carlo simulation results showing how the estimators based on $q = 1, q = q^*$ and $q = \hat{q}$ differ. Suppose that (i) the sequence $\{(x_t', u_t, v_t)\}$ is i.i.d. and (i) $f_t(0|x_t) = f(0)$. Then, the limiting distribution in Proposition 3 simplifies as follows:

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

where $\sigma_0^2(q) = E(\zeta_t^2), \zeta_t = qf(0)^{-1}\psi_\theta(v_t) + u_t^* - qv_t^*$ and $Q_{zz} = H(\Pi_0^*)' Q H(\Pi_0^*)$. Hence, in this case, the optimal choice for q can be obtained by minimising $\sigma_0^2(q)$ and it is given by

$$q^* = \frac{E(v_t^* u_t^*) - f(0)^{-1} E(\psi_\theta(v_t) u_t^*)}{f(0)^{-2}\theta(1-\theta) + E(v_t^{*2}) - 2f(0)^{-1}E(\psi_\theta(v_t) v_t^*)}. \tag{7}$$

Using the 'plug-in principle,' a consistent estimator for q^* is easily obtained based on any consistent kernel-estimator $\hat{f}(0)$ for $f(0)$;

$$\hat{q} = \frac{\sum_{t=1}^T \tilde{v}_t^* \tilde{u}_t^* - \hat{f}(0)^{-1} \sum_{t=1}^T \psi_\theta(\tilde{v}_t) \tilde{u}_t^*}{T.\hat{f}(0)^{-2}\theta(1-\theta) + \sum_{t=1}^T \tilde{v}_t^{*2} - 2\hat{f}(0)^{-1} \sum_{t=1}^T \psi_\theta(\tilde{v}_t) \tilde{v}_t^*} \tag{8}$$

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

6 Monte Carlo Simulations

We conduct simulation experiments to investigate the finite sample properties of the 2SQR(θ, q) in three cases: (i) the benchmark case ($q = 1$), (ii) when the optimal value ($q = q^*$) is used and finally (iii) when our consistent estimator ($q = \hat{q}$) is used.

The data generating process used in the simulations is described in Appendix B. The equation of interest is over-identified and the parameter values are $\beta_0' = (1, 0.2)$ and $\gamma_0 = 0.5$. 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 from a normal distribution at each of the 1,000 replications. For each replication, we estimate the parameter values β_0 and γ_0 and the deviations of the estimates from the true values. Then, we compute the sample mean and sample standard deviation of these deviations over the 1,000 replications. The optimal value q^* for different values of θ and for different error distributions is obtained by simulating the formula in (7) while $\hat{q}$ is calculated through (8).

The results for the 2SQR(θ, q) with N(0,1) errors are in Table 1. The cases $q = 1, q = q^*$ and $q = \hat{q}$ are respectively shown in Tables 1(a), 1(b) and 1(c). In all cases, as predicted by the results in Sections 4 and 5, the intercept estimate is systematically biased and the biases do not diminish as the sample size increases. On the other hand, the 2SQR(θ, q) provides unbiased estimates for the slope parameters (β_{10} and γ_0) for all choices of q and all values of θ. This outcome on the intercept and slope estimates also takes place for Tables 2 and 3 based on $t(3)$ and LN(0,1) distributions. Table 1(b) shows that using the optimal value q^* (whose simulated values are shown in the first row of the table) dramatically improves the accuracy of the 2SQR(θ, q) in comparison with the benchmark case in Table 1(a); the efficiency gain ranges from 14% up to 50% depending on the value of θ. Generally, the gain is larger for extreme quantiles ($\theta = 0.05$ and 0.95) than for the middle quantiles ($\theta = 0.25, 0.5$ and 0.75); specifically the ranges are 46%-50% for $\theta = 0.05$, 17%-24% for $\theta = 0.25$, 14%-22% for $\theta = 0.5$, 21%-23% for $\theta = 0.75$ and 45%-47% for $\theta = 0.95$.

Actually in empirical work, the true value of q^* is not known even though $\hat{q}$ will be close to q^* in large samples. Hence, it is interesting to investigate the use of $\hat{q}$ on the 2SQR(θ, q) in small samples. The results for normal errors are in Table 1(c) where we also provide simulation means and standard deviations of $\hat{q}$ for different values of T and θ. The table demonstrates that with sample size as low as $T = 50$, the use of $\hat{q}$ can result in substantial efficiency gains (35%-50% for $\theta = 0.05$, 17%-24% for $\theta = 0.25$, 14%-22% for $\theta = 0.5$, 21%-23% for $\theta = 0.75$ and 35%-47% for $\theta = 0.95$) as compared with the case $q = 1$ The estimation accuracy of $\hat{q}$ improves as the sample size increases to $T = 300$ and the efficiency gain becomes larger. In fact, with $T = 300$, it does not make any difference whether to use either $\hat{q}$ or q^*.

As expected when the errors are generated from $t(3)$ with fat tails, the standard deviations of the sampling distributions of the 2SQR(θ, q) are much larger than that obtained when the errors are normal. When $t(3)$ is used to generate the error terms (Table 2), even with q^* the percentage reductions in standard deviation are small for middle quantiles. However, with extreme quantiles, substantial reductions can be achieved (55%-63% for $\theta = 0.05$ and 57%-60% for $\theta = 0.95$). When $\hat{q}$ is used (Table 2(c)), there are cases where the standard deviation of the 2SQR(θ, q) increases by 0.08%-1.25% when $T = 50$ and $\theta = 0.5$. This is expected because in these cases there is no gain even with q^* and the estimation of $\hat{q}$ just adds noise to the process. As the sample size grows to 300, the negligible negative gains disappear and the performance of the 2SQR(θ, q) based on $\hat{q}$ is nearly identical to the one based on q^* despite the fact that the estimated values of $\hat{q}$ are not very close to q^*. It seems likely that the surface of the function $\sigma_0^2(q)$ is very flat around q^* in these cases.

Finally, we turn to the lognormal distribution case whose results are displayed in Tables 3(a)-3(c). The standard deviations rise even more in the this case, indicating that the 2SQR(θ, q) may be particularly sensitive to asymmetry of error distributions. When the true value q^* is employed, the efficiency gain is phenomenal, regardless the values of T and θ; the results do not vary much with T and large variance reductions are achieved for all quantiles. It is well known that when the distribution is skewed to the right, parameter estimation by quantile regression for large quantiles are generally very poor. The simulation results show that this is the precisely the case in which our method can generate the maximum efficiency gain. Table 3(c) shows that, when $\hat{q}$ is used, there can be a large efficiency loss (29%-56% as compared with results based on $q = 1$) for small sample size ($T = 50$) and for small quantile ($\theta = 0.05$), although the general case is of efficiency gains. However, as with the other error distribution cases, when the sample size increase to $T = 300$, the use of $\hat{q}$ delivers large efficiency gains for all quantiles

7 Conclusion

We analyse in this paper the transmission of the asymptotic bias in two-stage estimation procedures where the first stage is asymptotically biased. We exhibit the algebraic structure that describe the bias transmission in the asymptotic representation of the estimator. This enables us to show that even for general cases with endogenous variables, an asymptotic bias occurring only on the intercept of the first-stage estimation is integrally and exclusively transmitted to the intercept for the second-stage estimation.

To illustrate this issue, we fully develop the case of the two-stage quantile regression estimators with random regressors, dependent and non identically distributed error terms when the first stage is implemented with least-squares estimators. These results permit valid inferences in models estimated using quantile regressions, in which the possible endogeneity of some explanatory variables is treated via ancillary predictive equations.

Moreover, for the two-stage quantile regressions substantial variance reduction is obtained in the context of two-stage estimation by reformulating the dependent variable by using predictions from the reduced-form estimation. This approach alleviates a frequently mentioned disadvantage of quantile regressions, namely their small efficiency.

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

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

[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] 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.

[7] Bassett, G.W. and R.W. Koenker (1978), "Asymptotic Theory of Least Absolute Error Regression," *Journal of American Statistical Association*, 73, 618-622.

[8] Bassett, G.W. and R.W. Koenker (1986), "Strong Consistency of regression Quantiles and related Empirical Processes," *Econometric Theory*, 2, 191-201.

[9] Buchinsky, M. (1995), "Quantile Regression, Box-Cox transformation model, and the U.S. Wage Structure, 1963-1987," *Journal of Econometrics*, 65, 109-154.

[10] Buchinsky, M. (1998), "Recent Advances in Quantile Regression Models. A Practical Guideline for Empirical Research," *The Journal of Human Resources*, XXXIII, 1, 88-126.

[11] 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.

[12] 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.

[13] Engle, R.F. and S. Manganelli (2003), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," mimeo University of California-San Diego.

[14] 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.

[15] Granger, C.W.J. and C.-Y. Sin (2000), "Modelling the Absolute Returns of Different Stock Indices: Exploring the Forecastability of an Alternative Measure of Risk," *Journal of Forecasting*, 19, 277-298.

[16] 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.

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

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

[19] Jureckova, J. (1984), "Regression Quantiles and Trimmed Least Sqaures Estimator under a General Design," *Kibernetica*, 20, 345-357.

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

[21] Kim, T. and C. Muller (2003), "Two-Stage Quantile Regressions with a First-Stage of Quantile Regressions," mimeo University of Notthingham.

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

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

[24] Lipsitz, S.R., G.M. Fitzmaurice, G. Molenberghs and L.P. Zhao (1997), "Quantile regression Methods for longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immonodeficiency Virus," *Applied Statistics*, 46, 463-476.

[25] Machado, J.A.F. and J. Mata (2001). "Earning Functions in Portugal 1982-1994: Evidence from Quantile Regressions." *Empirical Economics,* 26, 115-34.

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

[27] Malinvaud, E. (1970), "The Consistency of Nonlinear Regressions," *Annals of Mathematical Statistics*, 41, 956-969.

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

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

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

[31] Newey, W.K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

[32] Newey, W.K., J.L. Powell, and F. Vella (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-603.

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

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

[37] Powell, J. (1985), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.

[38] Sakata, S. (2001), "Instrumental Variable Estimation Based on Mean Absolute Deviation," mimeo University of Michigan, February.

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

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

Appendix A: Proofs

Proof of Lemma 1: Using Assumptions 1, 3(ii) and 3(v) and the fact that the quantile influence function $\psi_\theta(\cdot)$ is bounded, Theorem II.8 in Andrews (1990) implies

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

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

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

where $F_t(\cdot|x_t)$ is the conditional cdf of v_t. Let $G(\lambda) = q^{-1}T^{-1}\sum_{t=1}^T F_t(\lambda|x_t)x_t x_t'\Delta$. 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 $E(M_T(\Delta)) - E(M_T(0)) = -E\{G'(\xi_{T,t})\} = -q^{-1}E\{T^{-1}\sum_{t=1}^T f_t(\xi_{T,t}|x_t)x_t x_t'\}\Delta$.

Let $Q_T = E\left[T^{-1}\sum_{t=1}^T f_t(\xi_{T,t}|x_t)x_t x_t'\right]$, $Q_{0T} = E\left[T^{-1}\sum_{t=1}^T f_t(0|x_t)x_t x_t'\right]$

and consider the $(i,j)^{th}$ element of $|Q_T - Q_{0T}|$, which is given by

$$\begin{aligned}
& |T^{-1}\sum_{t=1}^{T} E\left(\{f_t(\xi_{T,t}|x_t) - f_t(0|x_t)\}x_{ti}x_{tj}\right)| \\
\leq\ & T^{-1}\sum_{t=1}^{T} E\left(|f_t(\xi_{T,t}|x_t) - f_t(0|x_t)|\ |x_{ti}|\ |x_{tj}|\right) \\
\leq\ & L_0 T^{-1}\sum_{t=1}^{T} E\left(|\xi_{T,t}|\ \ |x_{ti}|\ |x_{tj}|\right)
\end{aligned}$$

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 condition in Assumption 3(ii). Next, we note that

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

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

$$\sup_{||\delta||\leq L} ||M_T(\hat{\Delta}_1(\delta)) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta)|| = o_p(1) \tag{9}$$

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

$$M_T(\hat{\Delta}_1(\hat{\Delta})) = o_p(1) \tag{10}$$

because $T^{1/2}H(\hat{\Pi})'M_T(\hat{\Delta}_1(\hat{\Delta})) = \left[\left.\frac{\partial S_T}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}\right]_-$. Here, $H(\hat{\Pi})$ is bounded in probability: $H(\hat{\Pi}) = O_p(1)$ by Assumption 2, and $\left[\left.\frac{\partial S_T}{\partial\alpha}\right|_{\alpha=\hat{\alpha}}\right]_-$ is $o_p(1)$ because it is the vector of left-hand-side partial derivatives of the objective function in (4), evaluated at the solution $\hat{\alpha}$.

The next step is to show that $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1)$. Using the same argument as in Lemma 5.2 of Jureckova (1977), it can be proven that (9) implies that for any $\epsilon > 0$, there exist $L > 0, \eta > 0$ and a positive integer T_0 such that

$$P\left(\min_{||\Delta||\geq L}\left\|M_T(\hat{\Delta}_1(\hat{\Delta}))\right\|\right) < \epsilon \tag{11}$$

for any $T > T_0$. Hence, if $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$ is not bounded in probability, then (11) implies that $M_T(\hat{\Delta}_1(\hat{\Delta})) \neq o_p(1)$, which contradicts (10). Therefore, we have

$$\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1). \tag{12}$$

Therefore, the results in (9), (10) and (12) imply

$$q^{-1}Q_0\hat{\Delta}_1(\hat{\Delta}) = M_T(0) + o_p(1). \tag{13}$$

By rearranging terms in (13), we have the asymptotic representation for the 2SQR(θ, q):

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

where B_α, Q_{zz}^*, and Π_0^* are defined in the proposition. QED.

Proof of Proposition 2: On the one hand, $Q_0 = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ where Q_1 is a $K \times 1$ matrix and Q_2 is a $K \times (K-1)$ matrix. On the other hand, $H(\Pi_0^*) = \begin{bmatrix} I_{K_1} & \Pi_0^* \\ 0_{K_2 \times K_1} & \end{bmatrix}$ and $R = [H(\Pi_0^*)'Q_0H(\Pi_0^*)]^{-1} H(\Pi_0^*)'$. We want to prove that $RQ_1 = \begin{bmatrix} 1 \\ 0_{(K_1+G-1)\times 1} \end{bmatrix}$. Let $A = RQ_0$ that is a $(G+K_1) \times K$ matrix. Since RQ_1 is the first column of A, we just need to show that the first column of A is composed of a one at the first line and of zeros elsewhere. Let a' be the first column of A. We have $AH(\Pi_0^*) = RQ_0H(\Pi_0^*) = I_{(G+K1)}$ by definition of R. It follows that the first column of $AH(\Pi_0^*)$ is a' due to the arrangement of elements in $H(\Pi_0^*)$ in Proposition 2, while the first column of $I_{(G+K_1)}$ is $(1, 0, \ldots, 0)'$. Hence, $RQ_1 = \begin{bmatrix} 1 \\ 0_{(K_1+G-1)\times 1} \end{bmatrix}$. QED.

Proof of Proposition 3: Replacing the asymptotic representation of the first stage 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 $M = R[I, -Q_0Q^{-1}]$ and $S_t = (q\psi_\theta(v_t), qv_t^* - u_t^*)' \otimes x_t$.

Note that since x_t', u_t, v_t are strong-mixing by assumption, and S_t is a measurable function of x_t', u_t, v_t, it follows that S_t is also strong-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:

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

which in turn implies that

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

where $D_T = M V_T M'$. QED.

Appendix B: Simulation Design

The structural system is given by $B \begin{bmatrix} y_t' \\ Y_t' \end{bmatrix} + \Gamma x_t' = U_t'$, where $\begin{bmatrix} y_t' \\ Y_t' \end{bmatrix}$ is a 2×1 vector of endogenous variables, x_t' is a 4×1 vector of exogenous variables with the first element set to one, U_t' is a 2×1 vector of errors, $B = \begin{bmatrix} 1 & -0.5 \\ -0.7 & 1 \end{bmatrix}$ and $\Gamma = \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & -0.5 \end{bmatrix}$. We are interested in the first equation of the system and the system is over-identified by the zero restrictions $\Gamma_{13} = \Gamma_{14} = \Gamma_{22} = 0$. Here, the parameters in (1) are $\gamma_0 = 0.5$ and $\beta_0' = (1, 0.2)$, X_1 is the first two columns in X and u is the first column in U. The above structural equation can be written as $\begin{bmatrix} y & Y \end{bmatrix} B' = -X\Gamma' + U$, which gives the following reduced form equations $\begin{bmatrix} y & Y \end{bmatrix} = X \begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} + \begin{bmatrix} v & V \end{bmatrix}$, where $\begin{bmatrix} \pi_0 & \Pi_0 \end{bmatrix} = -\Gamma' (B')^{-1}$ and $\begin{bmatrix} v & V \end{bmatrix} = U (B')^{-1}$. We obtain $\pi_0' = (2.3, 0.3, 0.3, -0.15)$ and $\Pi_0' = (2.6, 0.2, 0.6, -0.3)$.

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

Table 1(a). Simulation Means and Standard Deviations of $2SQR(\theta, q=1)$: 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 $2SQR(\theta, q=q^*)$: 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 $2SQR(\theta, q=\hat{q})$: 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 $2SQR(\theta, q=1)$: t(3).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-1.20	-0.36	0.01	0.45	1.36
		Std	6.94	1.56	1.08	0.99	1.04
	$\tilde{\beta}_1$	Mean	-0.01	-0.01	0.00	0.00	0.00
		Std	0.89	0.32	0.26	0.32	0.80
	$\tilde{\gamma}$	Mean	-0.03	-0.01	0.00	-0.02	-0.06
		Std	1.43	0.45	0.40	0.51	1.19
T = 300	$\tilde{\beta}_0$	Mean	-1.21	-0.37	0.03	0.40	1.21
		Std	2.09	0.57	0.38	0.33	0.33
	$\tilde{\beta}_1$	Mean	0.00	-0.01	0.00	-0.01	0.00
		Std	0.29	0.12	0.11	0.12	0.30
	$\tilde{\gamma}$	Mean	0.00	0.00	-0.01	0.00	0.00
		Std	0.41	0.16	0.14	0.17	0.42

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

		θ (q^*)	0.05 (-0.080)	0.25 (0.526)	0.50 (0.828)	0.75 (0.528)	0.95 (-0.080)
T = 50	$\tilde{\beta}_0$	Mean	0.88	0.03	0.02	0.11	-0.73
		Std	2.63	1.63	1.09	0.90	0.40
	$\tilde{\beta}_1$	Mean	-0.01	-0.01	-0.01	-0.01	0.00
		Std	0.33	0.30	0.27	0.30	0.33
	$\tilde{\gamma}$	Mean	-0.03	-0.02	0.00	-0.02	-0.02
		Std	0.54	0.49	0.41	0.47	0.48
T = 300	$\tilde{\beta}_0$	Mean	0.95	-0.01	0.03	0.04	-0.91
		Std	0.86	0.54	0.38	0.31	0.16
	$\tilde{\beta}_1$	Mean	-0.01	0.00	0.00	-0.01	0.00
		Std	0.13	0.11	0.10	0.12	0.13
	$\tilde{\gamma}$	Mean	-0.01	0.00	-0.01	-0.01	0.00
		Std	0.17	0.16	0.14	0.16	0.17

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

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	0.26	0.13	0.05	-0.04	-0.24
		Std	5.18	1.69	1.19	0.87	0.93
	$\tilde{\beta}_1$	Mean	-0.03	-0.01	-0.01	-0.01	-0.02
		Std	0.54	0.30	0.28	0.29	0.43
	$\tilde{\gamma}$	Mean	-0.01	-0.02	-0.02	-0.01	0.01
		Std	1.08	0.50	0.45	0.46	0.75
	$\hat{q}$	Mean	0.16	0.27	0.29	0.30	0.28
		Std	0.63	0.28	0.24	0.25	0.44
T = 300	$\tilde{\beta}_0$	Mean	0.89	0.04	0.03	-0.01	-0.85
		Std	0.93	0.55	0.38	0.32	0.28
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.13	0.11	0.11	0.12	0.13
	$\tilde{\gamma}$	Mean	-0.01	0.00	-0.01	-0.01	0.00
		Std	0.17	0.15	0.14	0.16	0.18
	$\hat{q}$	Mean	0.02	0.46	0.57	0.46	0.05
		Std	0.16	0.18	0.14	0.17	0.16

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

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	-0.68	-0.50	-0.26	0.24	2.89
		Std	0.25	0.18	0.17	0.57	8.74
	$\tilde{\beta}_1$	Mean	0.00	0.01	0.00	-0.01	-0.04
		Std	0.17	0.16	0.17	0.34	1.51
	$\tilde{\gamma}$	Mean	-0.03	-0.03	-0.01	0.03	0.22
		Std	0.25	0.23	0.26	0.48	2.07
T = 300	$\tilde{\beta}_0$	Mean	-0.73	-0.56	-0.32	0.17	1.94
		Std	0.09	0.07	0.07	0.22	3.54
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.01
		Std	0.06	0.06	0.07	0.14	0.59
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.03
		Std	0.09	0.09	0.09	0.18	0.83

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

		θ (q^*)	0.05 (0.413)	0.25 (0.524)	0.50 (0.526)	0.75 (0.230)	0.95 (-0.090)
T = 50	$\tilde{\beta}_0$	Mean	0.11	-0.02	0.00	-0.01	-1.31
		Std	0.10	0.08	0.11	0.34	1.39
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	-0.01	0.00
		Std	0.06	0.07	0.11	0.19	0.21
	$\tilde{\gamma}$	Mean	0.01	0.01	0.02	0.04	0.05
		Std	0.09	0.09	0.16	0.29	0.32
T = 300	$\tilde{\beta}_0$	Mean	0.12	-0.03	-0.02	-0.06	-1.57
		Std	0.03	0.03	0.04	0.13	0.52
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.02	0.02	0.04	0.08	0.09
	$\tilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.03	0.03	0.06	0.11	0.12

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

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\tilde{\beta}_0$	Mean	0.00	0.15	0.12	-0.08	-1.26
		Std	0.49	0.20	0.16	0.36	5.69
	$\tilde{\beta}_1$	Mean	0.02	0.01	0.01	0.01	0.00
		Std	0.22	0.13	0.13	0.21	0.71
	$\tilde{\gamma}$	Mean	-0.07	-0.04	-0.02	-0.01	-0.05
		Std	0.39	0.24	0.22	0.39	1.33
	$\hat{q}$	Mean	0.52	0.37	0.31	0.14	0.01
		Std	0.21	0.15	0.17	0.26	0.57
T = 300	$\tilde{\beta}_0$	Mean	-0.01	0.05	0.02	-0.10	-1.48
		Std	0.13	0.09	0.07	0.12	0.76
	$\tilde{\beta}_1$	Mean	0.00	0.00	0.00	0.01	0.00
		Std	0.02	0.03	0.05	0.07	0.12
	$\tilde{\gamma}$	Mean	-0.01	-0.01	-0.01	-0.01	0.00
		Std	0.05	0.05	0.06	0.11	0.19
	$\hat{q}$	Mean	0.51	0.46	0.46	0.21	-0.09
		Std	0.08	0.07	0.07	0.18	0.14