Inconsistency transmission and variance reduction in two-stage quantile regression

Tae-Hwan Kim, Christophe Muller

Inconsistency transmission and variance reduction in two-stage quantile regression

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

[a]School of Economics, Yonsei University, Seoul, Korea; [b]Aix-Marseille Univ, CNRS, EHESS, Ecole Centrale, AMSE, Marseille, France, Aix-en-Provence Cedex, France

ABSTRACT

In this paper, we propose a new variance reduction method for quantile regressions with endogeneity problems, for alpha-mixing or m-dependent covariates and error terms. First, we derive the asymp-totic distribution of two-stage quantile estimators based on the fit-ted-value approach under very general conditions. Second, we exhibit an inconsistency 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 tradeoff between an inconsistency confined to the intercept estimator and a reduction of the variance of the slope estimator. Monte Carlo simulation results show the fine performance of our approach. In particular, by com-bining quantile regressions with first-stage trimmed least-squares estimators, we obtain more accurate slope estimates than 2SLS, 2SLAD and other estimators for a broad set of distributions. Finally, we apply our method to food demand equations in Egypt.

KEYWORDS: Two-stage estimation; Variance reduction; Quantile regression; Asymptotic bias

JEL CODES: C13; C30

1. Introduction

This paper considers an innovative strategy to improve the efficiency of the estimation of some regression models, especially when there is some interest in the distribution of effects. Improving the efficiency of estimates under endogeneity has always been a challenge as many attempts to correct for endogeneity, e.g., by using instrumental variables, much increase the asymptotic and small sample variances of the estimators. We focus on the estimation of a structural equation using quantile regression. Since the seminal work by Bassett and Koenker (1978) and Koenker and Bassett (1978), the literature on quantile regression has grown rapidly. However, models estimated with this method often involve endogenous variables, which lead to inconsistent estimation.

These endogeneity issues can be dealt with different approaches all based on some kind of instrumental or identifying variables. However, the often relatively low accuracy of quantile regressions (e.g., as compared to OLS in cases close to normality) is generally further deteriorated by these procedures, due to the imperfect correlation of instruments with the endogenous independent variables.

CONTACT Christophe Muller, christophe.muller@univ-amu.fr, Aix-Marseille Univ, CNRS, EHESS, Ecole Centrale, AMSE, Marseille, France, 14, avenue Jules Ferry, Aix-en-Provence Cedex, 13621, France.

In this paper, we consider the 'fitted-value approach' of dealing with these endogeneity issues because it allows us to propose and exploit a new method of variance reduction for quantile regressions. This implies to consider two estimation stages: first, an ancillary estimation of the fitted-values for the endogenous regressors, based on a set of exogenous independent variables; and second, the final estimation of the structural equation in which the endogenous regressors are substituted with their fitted-values.

A concrete empirical example that we shall use in our practical illustration is that of food demand Engel curves from household survey data. The dependent variable is the budget share of a given consumption category, say food for example. The main independent variable is typically the household total consumption expenditure. In that case, the total expenditure variable is likely to be endogenous, firstly, because any measurement error in the dependent variable, for example due to omissions of food transactions during the collection, should also appear in the total expenditure variable. Secondly, other factors may simultaneously affect both kinds of expenditure variables. We show results for such an example in our application in Sec. 5.

In social sciences, other empirical contexts where two-stage IV estimation is convenient are for equations of individual agent outcomes or individual decisions, as they generally involve endogenous independent variables. Some sources of such endogene-ity are common data contamination of dependent and independent variables, unob-served heterogeneity correlated with both kind of variables, and possibly two-side causality issues. Consider for example an individual wage equation for a sample of workers, in which the dependent variable (y_t) is the logarithm of individual wage rate, while the two independent variables in this equation are a dummy variable for the industrial sector (x_{1t}) and the worker's education level (Y_t). The individual birth quar-ter (x_{2t}) is often used as an instrument for the education level that is typically sus-pected to be the sole endogenous independent variable in the structural model, as in Angrist and Krueger (1991).

In other sciences, any common unobserved factors simultaneously correlated with dependent and independent variables would generate the same kind of endogeneity issues. For example, in medical sciences, an unobserved climatic factor (say, rains in tropical areas) may simultaneously affect (1) health status through disease such as malaria due to the multi-plication of anopheles mosquitoes, and (2) transport costs due to flooded roads. In that case, understanding some changes in infant weights, which may reflect inefficient health care delivery in villages, may be hampered by endogeneity issues. In this context, examining low quantiles may be vital as they are typically related to measures of child malnutrition.

A leading issue with the instrumentation strategy in statistics to deal with endogeneity problems is that it generally much diminishes the accuracy of estimates. Then, improving estimation efficiency is crucial. An example of a method to improve efficiency in statistics is by 'shrinking' a given estimator towards a fixed point or a data-dependent point. In their seminal work, Stein (1955) and James and Stein (1960) showed the inadmissibility of the LS estimator and introduced shrinkage methods based on an "optimal trade-off" between consistency and efficiency. This research direction has now given birth to a large literature.[1] Other approaches combining estimators with

[1]For example: Green and Strawderman (1991), Judge and Mittelhammer (2004), Tong and Wang (2007), Liang et al.(2011), Raheem et al. (2012).

the aim of reducing variance are also used in the literature exploiting 'control variate' settings, e.g. Henderson et al. (2012) for combinations of contaminated measurements, or Cheng and Peng (2007) recently for local likelihood models.

In all these cases, the optimal combination weight of the considered estimation is determined by minimizing the asymptotic variance of the combined estimator.[2] However, this approach is unattractive for the above two-stage estimation because it either makes the entire estimator inconsistent, (i.e., all coefficients of the model are affected by inconsistencies with shrinking estimators), and/or it requires the estimation of the joint distribution of the two combined estimators.

In this paper, in order to raise the efficiency of two-stage quantile regression estimators, we build on Amemiya (1982) and Powell (1983) for two-stage LAD and we combine the dependent variable y_t with its fitted value $\hat{y}_t$ from a preliminary estimation, using a weight q, so that $qy_t + (1-q)\hat{y}_t$ can be used as the dependent variable in the final stage of a two-stage quantile regression. This method was never analyzed comprehensively in general random settings. In particular, its complete asymptotic properties have never been established, extensive finite-sample properties have not been studied, and the method has never been empirically applied. As a consequence, the performance of this approach was unknown, and we fill this gap. We also analyze how the combination weight q can be determined optimally and estimated consistently.

Then, we make several contributions. 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 an 'inconsistency transmission property' that characterizes the asymptotic representation of our estimator. We then use this property to facilitate the analyses of the link of the reduced-form and structural models, and to confine inconsistencies on the intercepts. Third, we propose estimators of optimal weights in redefined dependent variables in order to raise the efficiency of the slopes for two-stage quantile regressions.

We show theoretically, with Monte Carlo simulations, and empirically, that using the 'optimal' composite dependent variable can substantially enhance the efficiency of the slopes of the two-stage quantile estimator, notably when compared to 2SLS or 2SLAD.

Note that although the estimators of the coefficients of the independent variables are consistent, allowing for analyses of explanatory or correlation structures, the predictor of the quantile levels emerging from the regression is generally inconsistent because of the inconsistent estimator of the intercept. Therefore, even though the estimation results can be used for inferences about the effects of independent variables, they cannot be used for inferences about the estimated quantile levels.

The paper is organized as follows. Sec. 2 discusses the model and the two-stage quantile regression. In Sec. 3, we show the asymptotic properties of the two-stage quantile regression estimator. In Sec. 4, we present Monte Carlo simulation results. In Sec. 5, we apply our method to Engel curve estimation. Finally, Sec. 6 concludes. All technical proofs are collected in Appendix A.

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

2. The model and the two-Stage Quantile Regression Estimators

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

$$y_t = x_{1t}'\beta + Y_t'\gamma + u_t = z_t'\alpha + u_t, \tag{1}$$

for $t = 1, ..., T$ and 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, $z_t = [x_{1t}', Y_t']', \alpha = [\beta', \gamma']'$ 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. This crucial assumption will allow us to confine inconsistencies to an intercept parameter, often less interesting for analysts than the slope coefficients. We further assume that Y_t can be predicted from the exogenous variables using the following linear equation that is assumed to be correctly specified:

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

where $x_t' = [x_{1t}', x_{2t}']$ is a K row vector with $K = K_1 + K_2$, Π is a $K \times G$ matrix of unknown parameters and V_t' is a G row vector of unknown error terms. The DGP will be completed by Assumptions 2 (or 2′) and 3(iv) below. However, we first discuss the reduced form as the restriction on errors in the DGP will be more conveniently expressed in terms of the reduced-form errors.

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

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

for $t = 1, ..., T$ and where

$$\pi = H(\Pi)\alpha \text{ with } H(\Pi) = \left[\begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi\right] \quad \text{and } v_t = u_t + V_t'\gamma. \tag{4}$$

Eqs. (2) and (3) are the basis of the first-stage estimation that yields estimators $\hat{\pi}$ and $\hat{\Pi}$ respectively of π and Π. This is important for this paper because, first, we follow the fitted-value approach that requires a preliminary estimation of Π and, second, we use a reformulation of the dependent variable that requires a preliminary estimation of π.

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 effects.[3] 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. As mentioned in Blundell and Powell (2006) in relation to the fitted-value approach: "the reduced-form for y may be of interest if the values of instrumental variables are control variables for the policy maker." For example, social program targeting can be

[3]The literature on the structural approach for quantile regressions is abundant. See for example: Kemp (1999), MaCurdy and Timmins (2000), Abadie et al. (2002), Chen et al. (2003), Hong and Tamer (2003), Honore and Hu (2004), Chernozhukov and Hansen (2005, 2006, 2008a,b), Ma and Koenker (2006), Chernozhukov, Imbens and Newey (2007), Lee (2007), Sakata (2007).

improved by using individual living standard predictions based on exogenous correlates and descriptive quantile regressions, for example to better focus transfer schemes on the poor (Muller and Bibi 2010). We follow the fitted-value approach in this paper because it will allows us to develop a new variance reduction method, which is not possible with the structural approach.

The fitted-value approach has been used empirically to estimate homogeneous (or constant) quantile effects.[4] The theoretical foundation for the fitted-value approach was first laid by Amemiya (1982) and Powell (1983) who investigated the two-stage least-absolute-deviations estimator in a simple setting. Chen (1988) and Chen and Portnoy (1996) study 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. Kim and Muller (2004) use a similar approach with instead 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.

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 consider conditional quantile or conditional mean restrictions on v_t in the fitted-value approach.

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 $\psi_\theta(z) = \theta - 1_{[z \le 0]}$ and $1_{[.]}$ is the indicator function. If the orthogonality conditions, $E(z_t \psi_\theta(u_t)) = 0$, were satisfied, then the 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 is generally not consistent. This is the endogeneity problem that prevents us from using elementary quantile regressions.

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

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

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 Theorem 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.[5]

The reformulation of the dependent variable as $qy_t + (1-q)\hat{y}_t$ was originally suggested by Amemiya (1982) to improve efficiency under normality in two-stage

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

[5]Powell does not deal with variance reduction, while he considers the case of OLS in the first stage and LAD in the second stage.

estimation with $0<q<1$ and LAD second-stage estimators. In our setting, the case $q=1$ 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 tradeoff between two estimation methods. Chen and Portnoy (1996) use LS and Trimmed LS estimators in the first stage and quantile regression estimators in the second stage, relaxing the normality assumption in Amemiya, while they do not consider the reformulation of the dependent variable.

3. Asymptotic properties of the 2SQR estimator

3.1. Hypotheses

It is generally possible to employ consistent estimators in the first stage. However, in order to exploit later on a tradeoff between bias and efficiency, we allow in Assumption 1 for inconsistent first-stage estimation with bounded asymptotic bias terms. 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 and Π will be brought up later on.

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

Assumption 1 allows for primitive assumptions that will be developed for each special case of interest later on (Assumption 3′ for OLS, Assumption 3″ for Trimmed Least Squares), as explained in Subsection 3.4. In the current form, it preserves the possibility of extending our analysis to other first-stage estimators if wished. When the inconsistency terms B_π and B_Π are vanishing as $T \to \infty$, the first-stage estimators are consistent. A case of inconsistency 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 in $\hat{\pi}$ is inconsistent.

We now specify the hypotheses for the data generating process.

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

Assumption 2′. The sequence $\{(x_t', u_t, v_t)\}$ is m-dependent.

Assuming an α-mixing process is assuming an asymptotic independence property.[6] It corresponds to an unusual generality for the DGPs considered in the quantile regression literature. 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. Using α-mixing hypotheses not only avoids

[6]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 $\varepsilon>0$.

to have to decide the relevant value of m to assume for m -dependent processes, but also allows for some heterogeneity in the m, and for long term memory. Finally, it captures the idea that the dependence may be vanishing asymptotically. In that sense, it seems to be more plausible than m-dependence for many scientific or economic phenomena. For example, Chen, Hansen and Carrasco (2010) have shown that stationary diffusion that may approximate some stable physical processes, are α-mixing. In particular, in finance analyses, bond prices can be described with α -mixing processes.

To obtain the asymptotic distribution of the 2SQR(θ, q) estimator, we impose the following usual regularity conditions.

Assumption 3.

(i) $H(\Pi + B_\Pi)$ 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.
There exists a positive number C

(v) > 0 such that $E(\|x_t\|^3)<C<\infty$ for any t.

Assumption 3. (v′) x_t is a bounded random vector.

Assumption 3(i) is analogous to the usual identification condition for simultaneous equations models. The inconsistency term B_Π appears in the condition because the first-stage estimator converges towards $\Pi + B_\Pi$. Assumption 3(ii) simplifies the demonstration of convergence of some 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. In the case when OLS were used for estimating Π and α, Assumption 3(i) jointly with the first part of Assumption 3(iii) would ensure that $E[x_t Y_t] \neq 0$. When other estimators are used for Π, and when α is estimated using quantile regression in the second stage, identification results from Assumption 3(i) and the second part of Assumption 3(iii), as we shall discuss in the proof of Theorem 1. The last condition is akin to the one in the conventional IV approach since it is necessary for consistency and for the inversion of the relevant empirical process to establish asymptotic normality.

Assumption 3(iv) is the assumption that zero is the given θth -quantile of the conditional distribution of v_t. 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, which is used to derive the asymptotic representation in the $m-$dependent case of Assumption 2′. We also use it to limit 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.

However, if one wishes to extend the results to the $\alpha-$mixing case of Assumption 2, then Assumption 3(v) must be replaced by the stronger Assumption 3(v′).

Assumption 3(iv) is central to our fitted-value approach. As traditional in the 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 (1983), 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, interpreting the quantile restriction in Assumption 3(iv) may not appear to be straightforward, although it has been used for a long time in the literature. One justification for Assumption 3(iv) is when the reduced form is interesting in itself, for example for policy makers using it to assess the effect of exogenous correlates on the dependent variable, without involving any structural reasoning. A case in which Assumption 3(iv) is satisfied is when u_t and V_t are independent of the x_t, which implies that v_t is independent of the x_t by construction. Finally, if one considers the most general reduced-form model for the conditional mean, one would have instead $v_t \equiv Y_t - E(Y_t|x_t)$. In that case, one would have by construction $E(v_t|x_t) = 0$. As a consequence, replacing x_t in (3) with some high degree polynomials of the x_t may approximate $E(Y_t|x_t)$ well enough to allow the assumption that $v_t \equiv Y_t - x_t'\Pi$ satisfies approximately $E(v_t|x_t) = 0$. The same kind of approximation reasoning can be applied to $E(\psi_\theta(v_t)|x_t)$ in Assumption 3(iv). Obviously, such an approach would work only if enough exogenous variables and large samples are available.

Under Assumption 3(iv), only constant effect quantile models can be specified and estimated. In that case, the intercept estimates for different quantile indices describe the parallel shifts in the unconditional quantiles of the error term.

Despite being limited to constant effect quantile models, the fitted-value approach has several advantages worth mentioning. First, it is computationally inexpensive, even with a large number of endogenous and exogenous variables. Second and chiefly, considering the fitted-value approach provides us with an opportunity to develop a new powerful method of variance reduction. We now study the asymptotic properties of 2SQR(θ, q) in the next section.

3.2. The asymptotic representation

Combining Lemma 1 in Appendix B 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.

Theorem 1. Suppose that Assumptions 1-3 and 3(v′) hold, or that Assumptions 1, 2′ and 3 hold. Then, the asymptotic representation for the 2SQR(θ, q) estimator is:

$$T^{1/2}(\hat{\alpha}-\alpha-B_\alpha) = RT^{-1/2}\sum_{t=1}^{T} x_t q\psi_\theta(v_t) + (1-q)RQ_0T^{1/2}(\hat{\pi}-\pi-B_\pi) - RQ_0T^{1/2}\big(\hat{\Pi}-\Pi-B_\Pi\big)\gamma + o_p(1),$$

where

$$B_\alpha = RQ_0\{(1-q)B_\pi - B_\Pi\gamma\}, R = Q_{zz}^{*-1}H(\Pi^*)', Q_{zz}^* = H(\Pi^*)'Q_0H(\Pi^*) \text{ and } \Pi^* = \Pi + B_\Pi.$$

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 component in $\hat{\pi}$ may imply contradictions between some chosen restrictions on errors in the first and second stages and cause inconsistencies, which will be explained in detail in the next section. The formula of B_α is obtained as the value allowing $T^{1/2}(\hat{\alpha} - \alpha - B_\alpha) = O_p(1)$, 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 $\alpha + B_\alpha$ naturally follows from Theorem 1.

3.3. The inconsistency

In this section, we show how an inconsistency on the intercept of the first-stage estimator can be appropriately dealt with the 2SQR(θ, q). One reason why this result has not been exhibited in the literature 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 inconsistency is transmitted to the two-stage estimator. Our analysis is based on the algebraic structure of the asymptotic representation. 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 inconsistencies on the intercepts of the first-stage estimators affect exclusively the intercept of the two-stage estimator.

According to the asymptotic representation in Theorem 1, inconsistencies in $\hat{\pi}$ and in $\hat{\Pi}$ are transmitted to the 2SQR(θ, q) through the matrix RQ_0. This feature occurs for any two-stage estimation procedure that shares the same algebraic structure for the asymptotic representation.

To isolate the intercept of the first-stage estimators, we decompose matrix Q_0 and the first-stage estimators $\hat{\pi}$ and Π as follows. Let $Q_0 = [Q_1 \; Q_2]$ 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 let $\hat{\pi} - \pi = \begin{bmatrix} \hat{\pi}_{(1)} - \pi_{(1)} \\ \hat{\pi}_{(2)} - \pi_{(2)} \end{bmatrix}$, where $\hat{\pi}_{(1)}$ is the estimator of the constant coefficient. This yields $RQ_0(\hat{\pi} - \pi) = RQ_1(\hat{\pi}_{(1)} - \pi_{(1)}) + RQ_2(\hat{\pi}_{(2)} - \pi_{(2)})$, where the first term in the right-hand-side term is inconsistent, while the second term is consistent, by assumption. The contribution of $\hat{\Pi}$ can be similarly decomposed. The characterization of RQ_1 in the next theorem will be useful to exhibit some features of the inconsistency of $\hat{\alpha}$ in the case of consistent first-stage slope estimators.

Theorem 2. Given that $R = [H(\Pi^*)'Q_0H(\Pi^*)]^{-1}H(\Pi^*)'$ and $H(\cdot)$ is defined as in (4), we have

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

Theorem 2 implies that, in that case, the only coordinate of $\hat{\alpha}$ with a possible inconsistency is the intercept. Theorem 2 is useful because empirical researchers may often pay little attention to the intercept term, whereas the slope coefficients often carry clearer explanatory meaning.

Consider an example in which one performs the first-stage estimations of (2) and (3) by using LS. In that case, since under Assumption 3(iv), the zero mean condition on v_t cannot be satisfied, the intercept LS estimator cannot be consistent.[7] However, the remaining slope coefficients can still be consistently estimated if the covariates are exogenous. Depending on the distributional restrictions on V_t imposed by the researcher, she may also use an estimator for Π yielding an inconsistent intercept term, although this case may be little relevant in practice.

Clearly, the researcher can eliminate the inconsistency 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. We propose instead to choose $q \neq 1$ and to use first-stage estimators with inconsistent terms confined only to the intercept. In that case, the slope coefficients in the structural equation can be consistently estimated, while its asymptotic variance depends on q. This generates a tradeoff between the inconsistency in the intercept estimator (the first coordinate in $\hat{\alpha}$) and the efficiency of the slope estimator (the remaining coordinates in $\hat{\alpha}$). Our approach contrasts with the shrinking estimators for which all of coefficients estimators are made inconsistent to reduce their variance.

Let $v_t^* = v_t - F^{-1}_{v_t|x_t}(\theta)$. Then, $E(\psi_\theta(v_t^*)|x_t) = \theta - P[v_t^* \leq 0|x_t] = \theta - P[v_t \leq F^{-1}_{v_t|x_t}(\theta)|x_t] = \theta - \theta = 0$, which describe the conditional quantile restriction characterizing v_t^*. As a consequence, it is easy to obtain a reduced-form quantile regression restriction, provided one accepts the introduction in the regression of a possible nuisance inconsistency term $F^{-1}_{v_t|x_t}(\theta)$. This term may affect all coefficients of the model, whether it is linear in x_t, or nonlinear. Let us now assume that u_t and V_t are independent of $\widetilde{x}_t$, defined as x_t from which has been eliminated the constant variable. Since $v_t = u_t + V_t\gamma$, this independence restriction implies $F^{-1}_{v_t|x_t}(\theta) = F^{-1}_{v_t}(\theta)$, and the nuisance term is thus confined to the intercept. Then, according to Theorem 2, an inconsistency is generated exclusively on the intercept term of the structural model.

However, the restriction $F^{-1}_{v_t|x_t}(\theta) = F^{-1}_{v_t}(\theta)$ for all θ also implies constant effect in the quantile regressions of interest. Although such characterization of instrumental variables may be deemed to be strong by some authors, it is often the way instrumental variables are intuitively found by empiricists: variables that are not connected at all with the model error seen as a remainder of the explanation of the dependent variable given the effects of explanatory variables.

On the other hand, starting instead from $E(\psi_\theta(v_t)|x_t) = 0$ and assuming that v_t and V_t are both independent of $\widetilde{x}_t$, we obtain the 'structural' restriction $E(\psi_\theta(u_t^*)|x_t) = 0$, where $u_t^* = u_t - F^{-1}_{u_t}(\theta)$, for a structural model with the correct value of parameters,

[7]Unless the probability support of v_t does not include the interval between the mean and the quantile of interest.

except perhaps for a non-vanishing inconsistency term $F_{u_t}^{-1}(\theta)$ on the intercept. In that sense, under the independence assumption that $F_{v_{t|x_t}}^{-1}(\theta) = F_{v_t}^{-1}(\theta)$ for all θ, reduced-form quantile regressions and structural quantile regressions can be seen as emerging from akin restrictions, with perhaps the exception of the intercept.

3.4. Asymptotic normality and covariances with LS and trimmed-LS predictions

In this sub-section, we examine the case of (non-robust) LS estimation and (robust) trimmed-least-squares (TLS) estimation of π and Π in the first stage. 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 preserve some efficiency of the final two-stage quantile estimator, when errors are close to normal. 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 was studied in Kim and Muller (2004). In this special case, there is no inconsistency 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 effect quantile models, the conditional mean includes the coefficients of the model, as conditional quantiles do, and one may want to assume that x_t is mean-independent of v_t and V_t as in Assumption 3′ below. This is more restrictive than what was assumed for Theorems 1 and 2, because we now wish to confine the possible bias to the intercept exclusively. Moreover, such assumption is typical in LS estimation, which is our first stage do.

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 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 θth 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 $E(V_t^*|x_t) = 0$.

The reduced-form equations for Y_t and y_t in (2) and (3) can then be rewritten by reallocating the bias to the intercept coefficient as follows:

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

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

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

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

The inconsistency term B_π is generally non-zero for $q \neq 1$. In contrast, B_Π can be non-vanishing 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 $\widetilde{\Pi}$ and $\widetilde{\pi}$ be the first-stage LS estimators based on (6) and (7) respectively and let $\widetilde{\alpha}$ be the corresponding 2SQR(θ, q) estimator. The asymptotic representation for $\widetilde{\alpha}$ is shown in Appendix B.

Due to the characterization of B_α in Theorem 2, we have $B_\alpha = ((1-q)E(v_t) - E(V_t')\gamma, 0, ..., 0)'$. As seen before, the intercept estimator is inconsistent, while the slope estimators are not. To derive the asymptotic normality of $\widetilde{\alpha}$, 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 $V_T = var(T^{-1/2}\sum_{t=1}^{T} S_t)$ is positive definite for T sufficiently large, where $S_t = (q\psi_\theta(v_t), qv_t^* - u_t^*)' \otimes x_t$, $u_t^* = v_t^* - V_t^{*'}\gamma$ 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.

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

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

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

The asymptotic result in Theorem 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 minimizing the asymptotic covariance matrix shown in Theorem 3. There are many ways of minimizing a multi-dimensional covariance matrix. For example, one may minimize some norm of the matrix (e.g., the mean square error). One may also minimize the standard error for a given coefficient of interest in the structural model.

In some special cases (e.g., IID), the effect of q on D_T is concentrated in all these calculus in a scalar function that summarizes the contributions of all the error terms in this matrix. In that case, a unique optimal value q^* can be obtained. In the general case, q^* can also be made explicit when the MSE is minimized. Consistent preliminary estimation of q^* does not perturb the asymptotic properties of the 2SQR, which can be characterized as a MINPIN estimator (Andrews 1994, p. 2263), as long as a stochastic equicontinuity condition of the global empirical process is valid.

Let us now consider 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 Theorem 3 simplifies into $\sigma_0^2(q)Q_{zz}^{-1}$, where $\sigma_0^2(q) = E(\hat{\zeta}_t), \zeta_t = qf(\bar{0})^{-1}\ \psi_\theta(v_t) + u_t^* - qv_t^*$ and $Q_{zz} = H(\Pi^*)\ QH(\Pi^*)$. In this case, we can easily obtain the optimal weight q^* and its consist-ent estimator $\hat{q}$ in Lemma 2 in Appendix B.

To address robustness concerns, we now propose an alternative 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

Table 1(a). Simulated Bias Means and Standard Deviations of 2SQR*1* (θ, q = 1) : N(0,1).

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

$a_{ij} = I_{[i=j \text{ and } X_i'\hat{\pi}(\mu)<y_i<X_i'\hat{\pi}(1-\mu)]}$, $\hat{\pi}(\mu)$ is the quantile regression estimator centered on a given quantile μ chosen a priori. Chen and Portnoy (1996) provide the TLS Bahadur representation. Let $\breve{\alpha}$ 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) $E(v_t|x_t)-E(t\_v_t)-\mu[F_v^{-1}(\mu)+F_v^{-1}(1-\mu)]=0$, 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) $E(V_{it}|x_t)-E(t\_V_{it})-\mu[F_{V_i}^{-1}(\mu)+F_{V_i}^{-1}(1-\mu)]=0$, 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 $\Delta_{\widetilde{v}}$ and $\Delta_{\bar{V}_j}$ such that $E|x_{ti}\widetilde{v}_t^*|^3<\Delta_{\widetilde{v}}$ and $E|x_{ti}\widetilde{V}_{jt}^*|^3<\Delta_{\bar{V}_j}$, for all i, j and t, where

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

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

The Bahadur representation, obtained from Theorem 1, and the representation for TLS (Ruppert and Carroll 1980) can be used to obtain the asymptotic representation of $\breve{\alpha}$, which is shown in Appendix B. The non-vanishing inconsistency term is given by $(1-q)\{E(t\_v_t)+\mu[F_v^{-1}(\mu)+F_v^{-1}(1-\mu)]\}-\widetilde{B}_{\Pi,1}\gamma$, where $\widetilde{B}_{\Pi,1}$ is a G-row vector whose j^{th} element is given by $E(t\_V_{jt})+\mu[F_{V_j}^{-1}(\mu)+F_{V_j}^{-1}(1-\mu)]$. The asymptotic representation (24) in Appendix B together with Assumptions 3″ and 4′ delivers the following characterization of the asymptotic normality of $\breve{\alpha}$.

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

$$\widetilde{D}_T^{-1/2}T^{1/2}\left(\breve{\alpha}-\alpha-\widetilde{B}_\alpha\right)\xrightarrow{d}N(0,I),$$

Table 1(b). Simulated Bias Means and Standard Deviations of 2SQR1 (0, 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)
$T = 50$	$\widetilde{\beta}_0$	Mean	0.59	0.23	−0.01	−0.26	−0.62
		Std	1.19	0.89	0.71	0.54	0.36
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.19	0.19	0.18	0.18	0.19
	$\widetilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.27	0.26	0.26	0.26	0.27
$T = 300$	$\widetilde{\beta}_0$	Mean	0.72	0.29	−0.01	−0.31	−0.74
		Std	0.44	0.34	0.27	0.21	0.14
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.07	0.07	0.07	0.07	0.07
	$\widetilde{\gamma}$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.10	0.10	0.10	0.10	0.10

where $\widetilde{D}_T = M\widetilde{V}_T M'$ *and* $M = R[I, -Q_0 Q^{-1}]$. The corresponding formulae for the optimal weight q^* and its consistent estimator $\hat{q}$ are shown in Lemma 3 in Appendix B.

Let us say a few words about comparision with 2SLS from a theoretical point of view. With 2SLS, the Assumptions 1, 2 or 2', 3(i) can be kept the same. Assumptions 3(ii), (iii), (iv), (v), (v') can be weakened. In particular, it is not necessary to define a matrix Q_0 that is now replaced by matrix Q in the calculus. Assumption 3(iv) is replaced by Assumption 3′ (i), which centers the reduced-form error term in the mean instead of the θ^{th} quantile. In these conditions, there is no conflict between first-stage and second-stage semi-parametric restrictions, and no bias terms B_π and B_Π that are now zero, and therefore no point of examining a bias transmission property. The parameter q disappears from the asymptotic representation, and therefore there is no opportunity of variance reduction by choosing this parameter. The well-known asymptotic representation of the 2SLS, which we don't show, shows that the estimation is now not robust outliers in the errors from the first and the second stages. The well-known formula of the asymptotic variance-covariance matrix of the 2SLS is much simpler than that of the 2SQR(θ, q), in particular because now $M = R[I, -I]$ and the function ψ_θ is replaced by the identity. However, the ordering of the two variance-covariance matrices is ambiguous and depends on the subjacent distribution of the errors, as suggest the following Monte Carlo simulations. All these reflections extend to other common estimators. In the next section, we present Monte Carlo simulation results, which exhibit how much variance reduction can be achieved in finite samples.

4. Monte Carlo simulations

We base our simulations on a simultaneous equation system with two equations, which is presented in Appendix C along with the simulation set-up. The first equation, which is the 'structural' equation of interest, contains two endogenous variables and two exogenous variables including a constant. A total of four exogenous variables are present in the whole system.

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

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

		θ	0.05	0.25	0.50	0.75	0.95
$T = 50$	$\widetilde{\beta}_0$	Mean	0.22	0.15	−0.01	−0.20	−0.26
		Std	1.49	0.91	0.72	0.54	0.40
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.22	0.19	0.18	0.19	0.22
	$\widetilde{\gamma}$	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
$T = 300$	$\widetilde{\beta}_0$	Mean	0.62	0.25	−0.01	−0.27	−0.62
		Std	0.46	0.34	0.27	0.21	0.16
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.07	0.07	0.07	0.07	0.07
	$\widetilde{\gamma}$	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

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. The equation of interest is assumed to be over-identified and the parameter values are set to $\beta' = (\beta_1, \beta_2) = (1, 0.2)$ 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 from a normal distribution independently of the errors. 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 (22) or (25), while $\hat{q}$ is estimated through (23) or (26).[8]

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. We have conducted the same set of simulations for the case of heteroskedastic errors and have found that the results are qualitatively the same as in the iid case. We first comment 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. When there is regular convergence towards a non-zero value, the bias should increase or decrease with T, until the limit point. Here it increases. The standard errors decreasing with T illustrate the convergence. Besides, we checked with higher T that the estimate does not go to the infinity. Comparing results with q^* and $\hat{q}$ show that the slow convergence of the intercept estimator towards its asymptotic value is partly due to the slow convergence of $\hat{q}$ towards q^*. For the latter, the intercept estimates at $T = 50$ and $T = 100$ are closer.

On the other hand, the 2SQR1(θ, q) estimates for the slope parameters (β_1 and γ) are consistent 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

[8]See Appendix B.

Table 2(a). Simulated Bias Means and Standard Deviations of 2SQR*1* (θ, q = *1*) : t (3).

		θ	*0.05*	*0.25*	*0.50*	*0.75*	*0.95*
$T = 50$	$\widetilde{\beta}_0$	Mean	−1.07	−0.34	0.01	0.42	1.36
		Std	7.32	2.04	1.42	1.06	1.05
	$\widetilde{\beta}_1$	Mean	0.04	0.02	0.01	−0.01	−0.01
		Std	0.88	0.33	0.28	0.34	0.85
	$\widetilde{\gamma}$	Mean	−0.06	−0.02	−0.01	−0.01	−0.1
		Std	1.48	0.59	0.51	0.54	1.43
$T = 300$	$\widetilde{\beta}_0$	Mean	−1.18	−0.40	−0.02	0.37	1.20
		Std	2.17	0.59	0.40	0.33	0.33
	$\widetilde{\beta}_1$	Mean	0.02	0.00	0.00	0.00	0.00
		Std	0.29	0.11	0.10	0.12	0.30
	$\widetilde{\gamma}$	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). Simulated Bias Means and Standard Deviations of 2SQR1 (θ, 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)
$T = 50$	$\widetilde{\beta}_0$	Mean	0.90	0.01	0.02	0.10	−0.74
		Std	2.98	1.89	1.42	1.00	0.43
	$\widetilde{\beta}_1$	Mean	0.02	0.02	0.01	0.01	0.02
		Std	0.35	0.32	0.28	0.31	0.34
	$\widetilde{\gamma}$	Mean	−0.03	−0.02	−0.01	−0.02	−0.03
		Std	0.58	0.55	0.51	0.50	0.53
$T = 300$	$\widetilde{\beta}_0$	Mean	0.87	−0.06	−0.02	0.02	−0.92
		Std	0.94	0.56	0.40	0.32	0.16
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.12	0.11	0.10	0.11	0.12
	$\widetilde{\gamma}$	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). Simulated Bias Means and Standard Deviations of 2SQR1 (θ, q = $\hat{q}$): t (3).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\widetilde{\beta}_0$	Mean	−0.19	0.10	0.07	−0.01	−0.22
		Std	13.6	2.02	1.66	1.04	0.86
	$\widetilde{\beta}_1$	Mean	0.03	0.01	0.01	0.01	0.01
		Std	0.56	0.30	0.31	0.31	0.52
	$\widetilde{\gamma}$	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	$\widetilde{\beta}_0$	Mean	0.80	0.01	−0.01	−0.04	−0.87
		Std	1.02	0.57	0.40	0.33	0.29
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.12	0.11	0.10	0.11	0.13
	$\widetilde{\gamma}$	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). Simulated Bias Means and Standard Deviations of 2SQRI (θ, q = 1) : LN(0,I).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\widetilde{\beta}_0$	Mean	−0.50	−0.35	−0.10	0.37	2.00
		Std	1.43	1.26	1.56	2.06	3.00
	$\widetilde{\beta}_1$	Mean	0.02	0.02	0.02	0.01	0.03
		Std	0.21	0.21	0.30	0.48	1.73
	$\widetilde{\gamma}$	Mean	−0.05	−0.05	−0.05	−0.05	−0.14
		Std	0.35	0.33	0.47	0.88	2.64
T = 300	$\widetilde{\beta}_0$	Mean	−0.74	−0.57	−0.33	0.15	1.85
		Std	0.49	0.50	0.56	0.66	1.08
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.01	0.02
		Std	0.08	0.09	0.11	0.19	0.65
	$\widetilde{\gamma}$	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). Simulated Bias Means and Standard Deviations of 2SQRI (θ, q = q^*) : LN(0,I).

		θ (q *)	0.05 (I.0388)	0.25 (I.051)	0.50 (0.972)	0.75 (0.167)	0.95 (−0.146)
T = 50	$\widetilde{\beta}_0$	Mean	−0.56	−0.41	−0.08	0.18	−1.25
		Std	1.41	1.25	1.56	1.57	0.91
	$\widetilde{\beta}_1$	Mean	0.02	0.02	0.02	0.02	0.03
		Std	0.21	0.21	0.30	0.40	0.45
	$\widetilde{\gamma}$	Mean	−0.05	−0.04	−0.05	−0.06	−0.07
		Std	0.35	0.33	0.47	0.65	0.71
T = 300	$\widetilde{\beta}_0$	Mean	−0.79	−0.62	−0.31	−0.08	−1.42
		Std	0.49	0.50	0.56	0.54	0.32
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.08	0.09	0.11	0.16	0.18
	$\widetilde{\gamma}$	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). Simulated Bias Means and Standard Deviations of 2SQRI (θ, q = $\hat{q}$) : LN(0,I).

		θ	0.05	0.25	0.50	0.75	0.95
T = 50	$\widetilde{\beta}_0$	Mean	−0.39	0.04	0.18	0.21	−1.01
		Std	1.55	1.58	2.20	2.38	1.82
	$\widetilde{\beta}_1$	Mean	0.02	0.02	0.02	0.02	0.02
		Std	0.22	0.25	0.34	0.40	0.75
	$\widetilde{\gamma}$	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	$\widetilde{\beta}_0$	Mean	−0.68	−0.40	−0.23	−0.03	−1.37
		Std	0.50	0.52	0.56	0.55	0.37
	$\widetilde{\beta}_1$	Mean	0.00	0.00	0.00	0.00	0.00
		Std	0.09	0.09	0.11	0.16	0.18
	$\widetilde{\gamma}$	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			$\hat{\beta}_1$	0.19	0.20	0.04
			$\hat{\gamma}$	0.26	0.28	0.06
2SLS	LS	LS	$\hat{\beta}_1$	0.21	0.41	0.42
			$\hat{\gamma}$	0.30	0.78	0.70
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.05)$	$\hat{\beta}_1$	0.22	0.56	0.22
			$\hat{\gamma}$	0.33	2.74	0.37
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.25)$	$\hat{\beta}_1$	0.19	0.30	0.25
			$\hat{\gamma}$	0.26	0.58	0.41
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.50)$	$\hat{\beta}_1$	0.19	0.31	0.34
			$\hat{\gamma}$	0.26	0.59	0.66
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.75)$	$\hat{\beta}_1$	0.19	0.31	0.40
			$\hat{\gamma}$	0.27	0.52	1.01
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.95)$	$\hat{\beta}_1$	0.22	0.52	0.75
			$\hat{\gamma}$	0.32	1.19	2.06

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

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

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

Estimator	First Stage	Second Stage		N(0,l)	t (3)	LN(0,l)
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$(\theta=0.05)$	$\hat{\beta}_1$	0.07	0.12	0.09
			$\hat{\gamma}$	0.10	0.19	0.12
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.25)$	$\hat{\beta}_1$	0.07	0.11	0.09
			$\hat{\gamma}$	0.10	0.16	0.13
2SQR$(\theta,\hat{q})$	LS	Quantile$(\theta=0.50)$	$\hat{\beta}_1$	0.07	0.10	0.11
			$\hat{\gamma}$	0.10	0.15	0.16
2SQR$(\theta,\hat{q})$	LS	Quantile $(\theta=0.75)$	$\hat{\beta}_1$	0.07	0.11	0.16
			$\hat{\gamma}$	0.10	0.16	0.23
2SQR$(\theta,\hat{q})$	LS	Quantile $(\theta=0.95)$	$\hat{\beta}_1$	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,I)	t (3)	LN(0,I)
Trim(μ) -2SQR $(\theta, 1)$	Trim(0.25)	Quantile ($\theta = 0.05$)	$\hat{\beta}_1$	0.14	0.28	0.05
			$\hat{\gamma}$	0.20	0.42	0.07
Trim(μ) -2SQR $(\theta, 1)$	Trim(0.25)	Quantile $\theta = 0.25$)	$\hat{\beta}_1$	0.09	0.11	0.06
			$\hat{\gamma}$	0.13	0.16	0.08
Trim(μ) -2SQR $(\theta, 1)$	Trim(0.25)	Quantile ($\theta = 0.50$)	$\hat{\beta}_1$	0.09	0.09	0.09
			$\hat{\gamma}$	0.12	0.13	0.12
Trim(μ) -2SQR $(\theta, 1)$	Trim(0.25)	Quantile ($\theta = 0.75$)	$\hat{\beta}_1$	0.09	0.11	0.17
			$\hat{\gamma}$	0.13	0.16	0.24
Trim(μ) -2SQR $(\theta, 1)$	Trim(0.25)	Quantile ($\theta = 0.95$)	$\hat{\beta}_1$	0.13	0.30	0.62
			$\hat{\gamma}$	0.19	0.41	0.89
Trim(μ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile ($\theta = 0.05$)	$\hat{\beta}_1$	0.08	0.10	0.05
			$\hat{\gamma}$	0.11	0.14	0.07
Trim(μ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile ($\theta = 0.25$)	$\hat{\beta}_1$	0.08	0.09	0.06
			$\hat{\gamma}$	0.11	0.13	0.08
Trim(μ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile ($\theta = 0.50$)	$\hat{\beta}_1$	0.08	0.09	0.09
			$\hat{\gamma}$	0.11	0.13	0.12
Trim(μ)-2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile ($\theta = 0.75$)	$\hat{\beta}_1$	0.08	0.09	0.12
			$\hat{\gamma}$	0.11	0.13	0.17
Trim(μ) -2SQR $(\theta, \hat{q})$	Trim(0.25)	Quantile ($\theta = 0.95$)	$\hat{\beta}_1$	0.08	0.10	0.10
			$\hat{\gamma}$	0.11	0.14	0.14

Table 6. Summary Statistics: 2012 Egypt.

	mean	Std.
Food share	0.2074	0.0971
Total Expenditure (EL per year)	25,813	16,013
Earnings (EL per year)	29,963	21,051
Dummy for Kids	0.614	0.487

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. Indeed, this is what pushing $\hat{q}$ to zero does, as can be seen in (5). The efficiency 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, even though 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, say less than 0.5.

Table 7. Mean Regression of Food Share.

		Structural Equation	
	Intercept	Dummy for Kids	Log Total Expenditure
OLS	1.326 (0.0231)	0.0158 (0.0008)	−0.0938 (0.0023)
2SLS	1.2659 (0.0256)	0.0155 (0.0008)	−0.0878 (0.0026)
		Predictive Equation	
	Intercept	Dummy for Kids	Log Earnings
OLS	1.4532 (0.0472)	0.0270 (0.0018)	0.8405 (0.0046)
TLS	1.0571 (0.0958)	0.0203 (0.0037)	0.8830 (0.0094)
		Reduced-Form Equation	
	Intercept	Dummy for Kids	Log Earnings
OLS	1.1384 (0.0226)	0.0131 (0.0009)	−0.0738 (0.0022)
TLS	1.1622 (0.0451)	0.0133 (0.0017)	−0.764 (0.0044)

Standard errors are in parentheses. 7483 observations

Table 8. Quantile Regression of Food Share.

			Estimation Results		
Quantiles	q_hat		Intercept	Dummy for Kids	Log Total Expenditure
$\theta = 0.05$		*2SQR1*(θ,q =1)	1.1001	0.0218	−0.0881
			(0.0501)	(0.0017)	(0.0050)
	0.0498	*2SQR1*(θ,$q = 0\hat{q}$)	1.2576	0.0158	−0.0878
			(0.0255)	(0.0008)	(0.0026)
		Trim (μ)-2SQR *(θ,q =1)*	1.0608	0.0211	−0.0838
			(0.0477)	(0.0017)	(0.0048)
	0.0509	Trim (μ)-2SQR *(θ,$q = \hat{q}$)*	1.2439	0.0154	−0.0864
			(0.0244)	(0.0008)	(0.0024)
$\theta = 0.125$		2SQR1 (θ,q = 1)	1.2188	0.0184	−0.0905
			(0.0336)	(0.0011)	(0.0034)
	0.0908	*2SQR1(θ, $q = \hat{q}$)*	1.2616	0.0157	−0.0880
			(0.0255)	(0.0008)	(0.0026)
		Trim *(1u)*-2SQR (θ,q =1)	1.1783	0.0177	−0.0861
		–	(0.0321)	(0.0011)	(0.0032)
	0.0910	Trim (μ)-2SQR (θ,q=$\hat{q}$)	1.2468	0.0153	−0.0865
			(0.0244)	(0.0008)	(0.0024)
$\Theta = 0.50$		*2SQR1(θ,q = 1)*	1.3000	0.0156	−0.0916
			(0.0331)	(0.0011)	(0.0033)
	-0.0287	*2SQR1(θ,$q = \hat{q}$)*	1.2649	0.0155	−0.0877
			(0.0256)	(0.0008)	(0.0026)
		Trim (μ) *2SQR1(θ,q =1)*	1.2591	0.0149	−0.0872
			(0.0316)	(0.0011)	(0.0032)
	−0.290	Trim (μ) *2SQR1(θ,$q = \hat{q}$)*	1.2535	0.0150	−0.0865
			(0.0245	(0.0008	(0.0024)
$\Theta = 0.75$		*2SQR1(θ,q = 1)*	1.3541	0.0133	−0.0897
			(0.0381)	(0.0013)	(0.0038)
	−0.1059	*2SQR1(θ,$q = \hat{q}$)*	1.2709	0.0151	−0.0875
			(0.0255)	(0.0008)	(0.0025)
		Trim (μ) *2SQR1(θ,q =1)*	1.3139	0.0126	−0.0854
			(0.0364)	(0.00013)	(0.0036)
	−0.1065	Trim (μ) *2SQR1(θ,$q = \hat{q}$)*	1.2617	0.0418	−0.0866
			(0.0243)	(0.0008)	(0.0024)
$\Theta = 0.95$		*2SQR1(θ,q =1)*	1.3636	0.0073	−0.792
			(0.0553)	(0.0018)	(0.0055)
$\Theta = 0.95$		*2SQR1(θ,$q = \hat{q}$)*	1.2681	0.0154	−0.0878
			(0.0256)	(0.0009)	(0.0026)
		Trim (μ) *2SQR1(θ,q =1)*	1.3282	0.0067	−0754
			(0.0529)	(0.0018)	(0.0053)
	−0.0142	Trim (μ) *2SQR1(θ,$q = \hat{q}$)*	1.2564	0.0150	−0.0866
			(0.0245)	(0.0008)	(0.0024)

Given the substantial efficiency gains, it is natural to ask how close the reduced variance is to the Cramer-Rao (CR) lower bound. We have calculated the CR bound numerically for 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 $T = 50$. Since the intercept coefficient estimate is inconsistent, we only discuss the slope coefficients.

For a small sample size such as $T = 50$, the 2SLS efficiency loss is not negligible even for the normal distribution case, while it becomes 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 ($\theta = 0.25, 0.5$ 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 same middle quantiles, with almost a gain of a quarter of efficiency at the median. Finally, under lognormality, 2SLS performs badly relative to the CR bounds, while 2SQR1$(\theta, \hat{q})$ stays closer than 2SLS to the CR bounds for small and middle quantiles.

The results are almost the same when the sample size is increased to $T = 300$, as shown in Table 5(a). For the normal errors, both 2SLS and 2SQR1$(\theta, \hat{q})$ attain the CR bounds. However, the previous remarks still hold for the other two error distributions t(3) and LN(0,1). That is: (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.[9] According to our simulations, using a trimming threshold of $\mu = 0.25$ yields more accurate results than lower thresholds, such as 0.05 or 0.10.[10] 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 $q = \hat{q}$, can be compared. Clearly, our proposed variance reduction method works again very well with the 2SQR2 estimator.

The 2SQR2$(\theta, \hat{q})$ performs 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 due to sampling variations since this irregularity vanishes when $T = 300$. The improvement brought by 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 two most accurate 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 useful information. With Student errors, the 2SQR2$(\theta, \hat{q})$ is often the more accurate estimator, yielding results

[9] We have also tried LAD in the first-stage, as in Chen and Portnoy (1996). Since the results are almost identical to that of the 2SQR2, we do not show them. They are available upon request. Thus, what seems to matter here is the robustness of the first stage estimator.

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

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 produces the best accuracy. For upper quantiles, reformulating the dependent variables even 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 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.

5. An application to Engel curve estimation

In this section, we apply our proposed method to estimate a model of food demand Engel curves in Egypt. That is the explanation of household food budget shares by the total expenditure levels. The dataset is drawn from the 2012 Egypt HIECS survey for a sample of 7483 households. The model is similar to the one in Blundell, Chen, and Kristensen (2007). Linearizing these authors' specifications, we consider the following quantile Engel curve equation:

$$y_i = \beta_{0\theta} + \beta_{1\theta} x_{1i} + \gamma_\theta Y_i + u_i, \tag{8}$$

where y_i is the food budget share of household i in 1995, x_{1i} is a dummy variable for children (i.e., $x_{1i} = 0$ if household i has no children and 1 if household i has at least one child), and Y_i is the log of total expenditure on both nondurable goods and services of household i in 1995. Variable u_i is an error term subject to a conditional quantile restriction. Typically, Y_i may be endogenous, for example because of measurement errors in the expenditure data. Following Blundell, Chen, and Kristensen (2007) and Chen and Pouzo (2009), the male log-earnings of household i in 1995 is used as an instrument for Y_i. Table 6 shows descriptive statistics for the variables of the model.

In Table 7, we first present conditional mean estimates of the model, obtained by using OLS and 2SLS. The estimated coefficients of total expenditure are significantly negative, which indicates that, on average, food share decreases as total expenditure increases, as expected from the Engel law according to which the importance of food within consumption diminishes as household living standard increases. The presence of children positively affects the budget share devoted to food, probably partly as a consequence of lower living standards in larger families with identical incomes, and perhaps also because of different household preferences when children are present. Although the estimated coefficients of the dummy for children are similar for OLS and 2SLS, those of the total expenditure variables differ more, hinting at the presence of some endogeneity. The p-value of the usual Hausman test of exogeneity is almost zero, with a corresponding t-statistics of 29.55, which confirms the presence of endogeneity. The OLS and TLS estimates of the prediction equation that is used to generate the fitted-value, and the OLS and TLS estimates of the reduced form are also shown in Table 7.

The estimation results of our proposed methods are reported in Table 8 for a range of selected quantile indices $(\theta = 0.05, 0.25, 0.5, 0.75, 0.95)$. For each selected quantile

index θ, we estimate the quantile Engel curve in (8) using (i) two-stage quantile regression without the composite dependent variable, $2SQR1(\theta, 1)$, (ii) two-stage quantile regression with the composite dependent variable based on the optimally estimated combination weight, $2SQR1(\theta, \hat{q})$, (iii) two-stage quantile regression with TLS in the first stage and no reformulation of the dependent variable, $2SQR2(\theta, 1)$, (iv) two-stage quantile regression with TLS in the first stage and optimally reformulated dependent variable, $2SQR2(\theta, \breve{q})$. Since we are dealing with cases with approximately constant effect, the obtained estimated coefficients of log total expenditure and of the dummy for kids are close for all the estimation methods and across quantiles, and we concentrate our comments on variance reduction results. We do not discuss the standard errors for the estimates of the intercept coefficient, since they are deliberately inconsistent as a tradeoff in our variance reduction method.

As expected, with the $2SQR1(\theta, 1)$ standard errors are larger than with 2SLS. This is due to the well known accuracy loss in most cases when using quantile regression instead of LS. In contrast, when redefining optimally the dependent variable with $2SQR1(\theta, \hat{q})$, the standard errors across the different quantile indices are significantly reduced to a level comparable to the benchmark case of 2SLS. The estimates based on TLS, $2SQR2(\theta, 1)$, share the loss of accuracy of $2SQR1(\theta, 1)$ when the original dependent variable is used. However, a further gain in accuracy is obtained when redefining the dependent variable, i.e. with $2SQR2(\theta, \breve{q})$. In that case, slightly lower standard errors than with 2SLS can be achieved. The estimated standard errors of γ are: 0.0026 for 2SLS against 0.0024 for $2SQR2(\theta, \breve{q})$ for all tried θ. The gain in accuracy achieved with $2SQR2(\theta, \breve{q})$ yields a 12 percent reduction in standard errors for γ. For β_{10}, the efficiency gain is negligible.

Interestingly, for all quantiles, the estimated weights $\hat{q}$ and $\breve{q}$ are relatively small, ranging from -0.029 with TLS and $\theta = 0.5$ up to 0.091 with TLS and $\theta = 0.025$. This implies that the role of the fitted-value of y_t in the redefined dependent variable is predominant with these data at all quantile indices. Positive weights are found for quantiles 0.05 and 0.25, and negative ones for quantile indices 0.5, 0.75 and 0.9.

6. Conclusion

In this paper, we develop a new method of variance reduction for constant-effect two-stage quantile regression allowing for random regressors as well as for non-iid error terms.

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. This combination introduces a tradeoff between an inconsistency 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 the tradeoff, 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 this two-stage quantile regression estimator. Then, we apply our variance reduction method by choosing optimally the combination weight in the redefined dependent variable. Our Monte Carlo simulation results show massive efficiency gains in most 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 may yield more accurate or equivalent estimates than OLS, and second that it does not require the knowledge of the distribution shape, which is a shortcoming 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, the dependent variable should be reformulated 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 minimizing the MSE may be natural.

Let us recap the computation steps for the 'trimmed least-squares plus quantile regression' procedure: (1) trimmed least-squares for the reduced-form equation and the ancillary equations, (2) calculus of the fitted values for the endogenous regressors in the structural equation, (3) preliminary quantile regression of the structural equation in which 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 estimated weight, (7) final quantile regression of the structural equation.

It is worth mentioning that the proposed method for variance reduction may not be necessarily limited to two-stage quantile regression methods. In fact, our approach might be generalizable to other two-stage regression methods in which the use of a composite dependent variable does not disturb the consistency of the final slope estimator. However, finding these methods seems to be a challenge in itself. It seems that the empirical contexts and estimation frameworks one can think of for variance reduction via asymmetric bias transmission mostly fit situations where quantile regressions are used.

One may however examine the different steps of the calculus and the proofs to find out what the general principles behind the method are. First, any (regular) first-stage estimation method can be considered, as it is clear from our Assumption 1 and Theorem 1. Therefore, what matters for the generalization is the choice of the second-stage estimation method. Second, the bias transmission property, described in our Theorem 2, extends to any estimator defined by regular FOCs that can be inverted to obtain the asymptotic representation, as long as an intercept is present in the linear model. Third, a necessary ingredient is that the influence function of the considered second-stage estimator should preserve the corresponding semi-parametric restriction when the dependent variable is reformulated. When limiting the attention to reformulations involving affine transformations, conditional on exogenous variables, it seems that only two obvious functions satisfy this requirement: that of quantile regression and that of OLS. We conjecture that they are the only solutions under our hypotheses, although perhaps other influence functions may work for non-continuous supports or by restricting the family of admissible distributions. Finally, when these conditions are satisfied, the resulting asymptotic variance must depend on the weight parameter q, so that variance can be reduced through optimization with

respect to q. In the case of LS, the variance does not depend on q. It is therefore impossible to extend our variance reduction method to LS. As a consequence, at least in general settings, it seems that only quantile regressions can, so far, be employed for this method of variance reduction. Further progress may therefore involve additional hypotheses on the estimation problem to relax this limitation.

Acknowledgments

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, Durham University, 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.

Funding

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 support from the A*MIDEX project (No. ANR-11-IDEX-0001-02) funded by the" Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR).

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 $V^*_{Ti}(\zeta) = M^*_{Ti}(\zeta) - E(M^*_{Ti}(\zeta))$. Let us consider the case of α-mixing covariates and error terms in Assumption 2. 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 a bounded subset of R^K and for all $L_1>0$ in a neighborhood of zero,

$$\sup_{t\leq T, T>1}\left[E\left(\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m^*_i(w_t,\zeta_1)-m^*_i(w_t,\zeta)|^p\right)\right]^{1/p} \leq CL_1^{\psi}, \tag{9}$$

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

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

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

$$\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1} |m_i^*(w_t,\zeta_1)-m_i^*(w_t,\zeta)|^2$$
$$\le x_{ti}^2 \sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1} 1_{\left[|qv_t-x_t'\zeta|\le\|x_t\|\times\|\zeta_1-\zeta\|\right]} \le x_{ti}^2 1_{\left[|qv_t-x_t'\zeta|\le\|x_t\|L_1\right]},$$

which implies

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

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,

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

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 $\psi=1/2$,

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

Hence, condition (b) is a consequence of Assumption 1. Therefore, by Theorem II.8 in Andrews (1990), $\{V_{T_i}^*(\zeta):T\ge1\}$ 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

$$\sup_{\|\zeta_1-\zeta_2\|\le L_T^*}\|V_T^*(\zeta_1)-V_T^*(\zeta_2)\|=o_p(1). \tag{10}$$

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 $K\times1$ 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 (10) becomes

$$\sup_{\|\Delta_1-\Delta_2\|\le L}\|V_T(\Delta_1)-V_T(\Delta_2)\|=o_p(1). \tag{11}$$

Setting $\Delta_1=\Delta$ and $\Delta_2=0$ in (11), yields

$$\sup_{\|\Delta\|<L}\|M_T(\Delta)-M_T(0)-\{EM_T(\Delta)-EM_T(0)\}\|=o_p(1). \tag{12}$$

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

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

Therefore, we have

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

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 $E(M_T(\Delta))-E(M_T(0)) = -E\{G'(\xi_{T,t})\} = -q^{-1}E\{T^{-1}\sum_{t=1}^{T} f_t(\xi_{T,t}|x_t)x_t x_t'\}\Delta$. We now examine the convergence of this term.

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

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

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

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

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 $E(M_T(\Delta))-E(M_T(0)) \to -q^{-1}Q_0\Delta$. The case of m-dependent covariates and error terms can be dealt with similarly by using theorems 1-3 in Andrews (1994) as in his example on page 2273, instead of Theorem II.8 in Andrews (1990). QED.

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

$$
M_T\big(\hat{\Delta}_0\big) = M_T(0) - q^{-1}Q_0\hat{\Delta}_0 + o_p(1), \tag{13}
$$

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 $q>0$.[11] 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

$$
M_T\big(\hat{\Delta}_0\big) = O_p(1). \tag{14}
$$

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

$$
\sup_{||\delta||\le L} ||M_T\big(\hat{\Delta}_1(\delta)\big) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta)|| = o_p(1). \tag{15}
$$

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

[11] 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).

We now use the argument between (A.7) and (A.8) in Powell (1983) to show that (14) and (15) imply that for some finite $L_2>0$:

$$\sup_{||\delta||\le L_2} ||\widetilde{M}_T(\delta)-H(\Pi^*)'M_T\big(\hat{\Delta}_0\big)+q^{-1}Q^*_{zz}\delta|| = o_p(1), \tag{16}$$

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

$$\begin{aligned}
||\widetilde{M}_T(\delta)-H(\Pi^*)'M_T\big(\hat{\Delta}_0\big)+Q^*_{zz}\delta|| \le\ & ||H(\hat{\Pi})||\ ||M_T\big(\hat{\Delta}_1(\delta)\big)-M_T(0)+q^{-1}Q_0\hat{\Delta}_1(\delta)|| \\
&+||H(\hat{\Pi})-H(\Pi^*)||\ ||M_T(0)-q^{-1}Q_0\hat{\Delta}_0|| \\
&+||H(\hat{\Pi})-H(\Pi^*)||\ \Big\{||H(\hat{\Pi})||\ +||H(\Pi^*)||\Big\}||q^{-1}Q_0||||\delta||+||H(\Pi^*)||,
\end{aligned}$$

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

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

$$\widetilde{M}_T(\hat{\Delta})=o_p(1). \tag{17}$$

Note that

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

where

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

First, we have that $\hat{A}_t=0$ because $x_t'H(\hat{\Pi})\alpha=x_{1t}'\beta+x_t'\hat{\Pi}\gamma$ and $x_t'H(\Pi)\alpha=x_{1t}'\beta+x_t'\Pi\gamma$. 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 $T^{1/2}\widetilde{M}_T(\hat{\Delta})=\Big[\frac{\partial S_T}{\partial\alpha}|_{\alpha=\hat{\alpha}}\Big]_-$, 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 (17); i.e., $\widetilde{M}_T(\hat{\Delta})=o_p(1)$.

Next, let us show that $\hat{\Delta}=T^{1/2}(\hat{\alpha}-\alpha-B_\alpha)=O_p(1)$. 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) $-\delta'\widetilde{M}_T(\lambda\delta)\ge-\delta'\widetilde{M}_T(\delta)$ for $\lambda\ge1$ and $||\delta||\le L_3$ for some $L_3>0$,
(ii) $||H(\Pi^*)'M_T(\hat{\Delta})||=O_p(1)$,
(iii) $\widetilde{M}_T(\hat{\Delta})=o_p(1)$,
(iv) Q^*_{zz} is positive definite.

Condition (i) is obtained by noticing that function $h(\lambda)=\sum_{t=1}^{T}\rho_\theta(qv_t-T^{-1/2}x_t'\hat{\Delta}(\hat{\Pi})\delta\lambda)-\rho_\theta(qv_t-T^{-1/2}x_t'\hat{\Delta}(\hat{\Pi})\delta)$ is convex in λ, and therefore that its gradient, $\delta'\widetilde{M}_T(\lambda\delta)$ is non-decreasing in λ. Condition (ii) comes from (14). 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

$$\hat{\Delta} = T^{1/2}(\hat{\alpha}-\alpha-B_\alpha) = O_p(1). \tag{18}$$

Therefore, we can plug $\hat{\Delta}$ into (16) in place of δ to obtain the following result:

$$\widetilde{M}_T(\hat{\Delta})-H(\Pi^*)'M_T(\hat{\Delta}) + q^{-1}Q^*_{zz}\hat{\Delta} = o_p(1). \tag{19}$$

Note that the first term in (19) is $o_p(1)$ because of (17). Hence, we have

$$q^{-1}Q^*_{zz}\hat{\Delta} = H(\Pi^*)'M_T(\hat{\Delta}) + o_p(1) \tag{20}$$

$$= H(\Pi^*)'M_T(0)-H(\Pi^*)'q^{-1}Q\hat{\Delta} + o_p(1), \tag{21}$$

where the second equality comes from (13). Since we are now facing a generalized Taylor expansion of the FOCs for our estimator, it is easy to see that the analog of the usual identification condition corresponds here to stating that the factor $q^{-1}Q^*_{zz}$ of $\hat{\Delta}$ in this expansion can be inverted. This is equivalent to imposing that first $H(\Pi^*)$ is full rank, and second that Q_0 is finite and positive definite. These conditions are satisfied respectively from Assumptions A3(i) and A3(iii). By plugging the definition of $\hat{\Delta}$ into (21) and inverting $q^{-1}Q^*_{zz}$, we obtain

$$T^{1/2}(\hat{\alpha}-\alpha-B_\alpha) = Q^{*-1}_{zz}H(\Pi^*)'\left\{T^{-1/2}\sum_{t=1}^{T}x_t q\psi_\theta(v_t) + (1-q)Q_0T^{1/2}(\hat{\pi}-\pi-B_\pi)-Q_0T^{1/2}\left(\hat{\Pi}-\Pi-B_\Pi\right)\gamma\right\} + o_p(1),$$

which completes the proof. QED.

Proof of Theorem 2: On the one hand, $Q_0 = [\,Q_1 \quad Q_2\,]$ where Q_1 is a $K \times 1$ matrix and Q_2 is a $K \times (K-1)$ matrix. On the other hand, $H(\Pi^*) = \begin{bmatrix} I_{K_1} & \Pi^* \\ 0_{K_2\times K_1} & \end{bmatrix}$ and $R = [H(\Pi^*)'Q_0H(\Pi^*)]^{-1}H(\Pi^*)'$. We want to prove $RQ_1 = \begin{bmatrix}1\\ 0_{(K_1+G-1)\times 1}\end{bmatrix}$. Let $A = RQ_0$, a $(G+K_1)\times K$ matrix. Since RQ_1 is the first column of A, which we denote a', we just need to show that a' is composed of a one at the first line and zeros elsewhere. We have $AH(\Pi^*) = RQ_0H(\Pi^*) = I_{(G+K1)}$ by definition of R. It follows that the first column of $AH(\Pi^*)$ is a' owing to the arrangement of elements in $H(\Pi^*)$ in Theorem 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 Lemma 2: With OLS first-stage estimators, we have $\sigma^2_0(q) = aq^2 + 2bq + c$, where $a = E\left[\left(f(0)^{-1}\psi_\theta(v_t) - v^*_t\right)^2\right]$, $b = E\left[\left(f(0)^{-1}\psi_\theta(v_t) - v^*_t\right)u^*_t\right]$ and $c = E(u^{*2}_t)$, which corresponds to a convex parabolic curve that attains its minimum at

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

which completes the proof. QED.

Proofof 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 $\widetilde{v}_t = F^{-1}_v(\mu)1_{[v_t<F^{-1}_v(\mu)]} + F^{-1}_v(1-\mu)1_{[v_t<F^{-1}_v(1-\mu)]} + v_t 1_{[F^{-1}_v(\mu)<v_t<F^{-1}_v(1-\mu)]}$. 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 $E(\widetilde{v}_t \mid x_{(j)t}, j = 2, ...) = 0$ that has not been canceled, i.e. $E(\widetilde{v}_t)$. 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 $\widetilde{v}_t$, we have

$$E\widetilde{v}_t = F_v^{-1}(\mu)P\Big[v_t<F_v^{-1}(\mu)\Big] + F_v^{-1}(1-\mu)P\big[v_t<F_v^{-1}(1-\mu)\big] + \int_{F_v^{-1}(\mu)}^{F_v^{-1}(1-\mu)} vf_v(v)dv$$

$= F_v^{-1}(\mu)F[F_v^{-1}(\mu)] + F_v^{-1}(1-\mu)\{1-F[v_t<F_v^{-1}(1-\mu)]\} + E(t\_v_t)$, which gives the result:
$[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)]\mu + E(t\_v_t)$. Using these results, the proof of Lemma 3 is similar to that of Lemma 2. QED.

Appendix B: Technical details

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

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

where Δ is a $K \times 1$ vector. Applying Theorem II.8 in Andrews (1990) or Theorems 1-3 in Andrews (1994) yields the following lemma.

Lemma 1. Suppose that Assumptions 1-3 and 3(v′) hold, or that Assumptions 1, 2′ and 3 hold. Then, for any $L > 0$, we have the following result:

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

Asymptotic representation for $\widetilde{\alpha}$: The asymptotic representations of $\widetilde{\Pi}$ and $\widetilde{\pi}$ are obtained and plugged into the formula in Theorem 1 to obtain:

$$T^{1/2}(\widetilde{\alpha} - \alpha - B_\alpha) = RT^{-1/2}\sum_{t=1}^{T} x_t q\psi_\theta(v_t) - RQ_0Q^{-1}T^{-1/2}\sum_{t=1}^{T} x_t\big(qv_t^* - u_t^*\big) + o_p(1).$$

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 $\widetilde{\alpha}$ is given by:

$$q^* = \frac{E(v_t^*u_t^*) - f(0)^{-1}E\big(\psi_\theta(v_t)u_t^*\big)}{f(0)^{-2}\theta(1-\theta) + E\big(v_t^{*2}\big) - 2f(0)^{-1}E\big(\psi_\theta(v_t)v_t^*\big)}. \tag{22}$$

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

$$\hat{q} = \frac{\sum_{t=1}^{T}\hat{v}_t^*\hat{u}_t^* - \hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\hat{u}_t^*}{T\hat{f}(0)^{-2}\theta(1-\theta) + \sum_{t=1}^{T}\hat{v}_t^{*2} - 2\hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\hat{v}_t^*}, \tag{23}$$

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

Asymptotic representation for $\check{\alpha}$:

$$T^{1/2}\left(\check{\alpha}-\alpha-\widetilde{B}_{\alpha}\right)=RT^{-1/2}\sum_{t=1}^{T}x_t q\psi_\theta(v_t)-RQ_0Q^{-1}T^{-1/2}\sum_{t=1}^{T}x_t\left(q\widetilde{v}_t^*-\widetilde{u}_t^*\right)+o_p(1), \tag{24}$$

where the asymptotic bias vector $\widetilde{B}_{\alpha}$ is such that the first element only is non-zero.

Lemma 3. 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 $\widetilde{\sigma}_0^2(q)\widetilde{Q}_{zz}^{-1}$, where $\widetilde{\sigma}_0^2(q)=E(\widetilde{\zeta}_t^2)$, $\widetilde{\zeta}_t=qf(0)^{-1}\psi_\theta(v_t)+\widetilde{u}_t^*-q\widetilde{v}_t^*$ and $\widetilde{Q}_{zz}=H(\widetilde{\Pi}^*)'Q_0H(\widetilde{\Pi}^*)$, and $\widetilde{\Pi}^*=\Pi+\widetilde{B}_{\Pi}$. In this case, the optimal value of q minimizing $\widetilde{\sigma}_0^2(q)$ is given by:

$$q^*=\frac{E\left(\widetilde{v}_t^*\widetilde{u}_t^*\right)-f(0)^{-1}E\left(\psi_\theta(v_t)\widetilde{u}_t^*\right)}{f(0)^{-2}\theta(1-\theta)+E\left(\widetilde{v}_t^{*2}\right)-2f(0)^{-1}E\left(\psi_\theta(v_t)\widetilde{v}_t^*\right)}. \tag{25}$$

A consistent estimator for q^* is:

$$\hat{q}=\frac{\sum_{t=1}^{T}\check{v}_t*\check{u}_t*-\hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\check{u}_t*}{T\hat{f}(0)^{-2}\theta(1-\theta)+\sum_{t=1}^{T}\check{v}_t*2-2\hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\check{v}_t*}, \tag{26}$$

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 $\hat{\pi}_\theta=\operatorname{argmin}_\pi\sum_{t=1}^{T}\rho_\theta(y_t-x_t'\pi)$.

Appendix C: Simulation set-up

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.[12] 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

[12]Since we are looking at models with constant slopes not depending on the chosen quantile it is tantamount to regenerate the errors to satisfy the condition for each quantile, or to generate them for a given quantile and estimating the different condtional quantiles with the same data. The centering of the error tem in each quantile θ is merely a normalization of the intercept in this case.

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 $[\pi \quad \Pi] = -\Gamma'(B')^{-1}$, we generate the endogenous variables $[y \quad Y]$ by using the reduced-form equation: $[y \quad Y] = X[\pi \quad \Pi] + [v \quad V]$.

ORCID

Christophe Muller http://orcid.org/0000-0001-9763-1527

References

Abadie, A., J. Angrist, and G. Imbens. 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70 (1):91–117.

Amemiya, T. 1982. Two stage least absolute deviations estimators. *Econometrica* 50 (3):689–711.

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.

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.

Angrist, J. D., and A. B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106 (4):979–1014.

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 (1):7–40.

Bassett, G., and R. Koenker. 1978. Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association* 73 (363):618–22.

Blundell, R., X. Chen, and D. Kristensen. 2007. Semi-nonparametric IV estimation of shape invariant Engel curves. *Econometrica* 75 (6):1613–69.

Blundell, R., and J. L. Powell. 2006. "Endogeneity in Nonparametric and Semiparametric Regression Models, in *Advances in economics and econometrics: Theory and applications, eighth world congress*. M. Dewatripont, L.P. Hansen and S.J. Turnovsky ed. Cambridge: Cambridge University Press.

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.

Chen, L.-A., and S. Portnoy. 1996. Two-Stage Regression quantiles and two-Stage Trimmed least squares estimators for structural equation models. *Communication in Statistics* 25 (5):1005–32.

Chen, X., L. P. Hansen, and M. Carrasco. 2010. Nonlinearity and temporal dependence. *Journal of Econometrics* 155 (2):155–69.

Chen, X., O. Linton, and I. Van Keilegom. 2003. Estimation of semi-parametric models when the criterion function is not smooth. *Econometrica* 71 (5):1591–608.

Chen, X., and D. Pouzo. 2009. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics* 152 (1):46–60.

Cheng, M.-Y., and L. Peng. 2007. Variance reduction in multiparameter likelihood models. *Journal of the American Statistical Association, March* 102 (477):293–304.

Chernozhukov, V., and C. Hansen. 2005. An IV model of quantile treatment effects. *Econometrica* 73 (1):245–61.

Chernozhukov, V., and C. Hansen. 2006. Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132 (2):491–525.

Chernozhukov, V., and C. Hansen. 2008. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics* 142 (1):379–98.
Chernozhukov, V., and C. Hansen. 2008. The reduced form: A simple approach to inference with weak instruments. *Economics Letters* 100 (1):68–71.
Chernozhukov, V., G. W. Imbens, and W. K. Newey. 2007. Instrumental variable estimation of nonseparable models. *Journal of Econometrics* 139 (1):4–14.
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 (8):1705–23.
Chortareas, G., G. Magonis, and T. Panagiotidis. 2012. The asymmetry of the new Keynesian Phillips curve in the euro area. *Economics Letters* 114 (2):161–3.
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 (1):149–67.
Green, E. J., and W. E. Strawderman. 1991. James-Stein type estimator for combining unbiased and possibly biased estimators. *Journal of the American Statistical Association* 86 (416):1001–6.
Henderson, J. V., A. Storeygard, and D. N. Weil. 2012. Measuring economic growth from outer space. *The American Economic Review* 102 (2):994–1028.
Hjort, N. L., and D. Pollard. 1999. Asymptotics for minimisers of convex processes. Discussion paper, Yale University.
Hong, H., and E. Tamer. 2003. Inference in censored models with endogenous regressors. *Econometrica* 71 (3):905–32.
Honore, B. E., and L. Hu. 2004. On the performance of some robust instrumental variables estimators. *Journal of Business and Economic Statistics* 22 (1):30–9.
James, W., and C. Stein. 1960. Estimation with quadratic loss. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*. (vol.1), Berkeley, CA: University of California Press.
Judge, G. G., and R. C. Mittelhammer. 2004. A semiparametric basis for combining estimation problems under quadratic loss. *Journal of the American Statistical Association* 99 (466):479–87. Vol.
Jureckova, J. 1977. Asymptotic relations of M-Estimates And R-Estimates In linear regression model. *The Annals of Statistics* 5 (3):464–72.
Kemp, G. C. R. 1999. Least absolute error difference estimation of a single equation from a simultaneous equations system. Discussion paper, University of Essex.
Kim, T., and C. Muller. 2004. Two-stage quantile regressions when the first stage is based on quantile regressions. *The Econometrics Journal* 7:18–46.
Kim, T.-H., and H. White. 2003. Estimation, inference, and specification testing for possibly misspecified quantile regression. *Advances in Econometrics* 17 :107–32.
Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica* 46 (1):33–50.
Koenker, R., and Q. Zhao. 1996. Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12 (05):793–813.
Lee, S. 2007. Endogeneity in quantile regression models: a control function approach. *Journal of Econometrics* 141 (2):1131–58.
Liang, H., G. Zou, A. T. K. Wan, and X. Zhang. 2011. Optimal weight choice for frequentist model average estimators. *Journal of the American Statistical Association* 106 (495):1053–66.
Ma, L., and R. Koenker. 2006. Quantile regression methods for recursive structural equation models. *Journal of Econometrics* 134 (2):471–506.
MaCurdy, T., and C. Timmins. 2000. Bounding the influence of attrition on intertemporal wage variation in the NLSY. mimeo Stanford University.
Muller, C., and S. Bibi. 2010. Refining targeting against poverty: Evidence from Tunisia. *Oxford Bulletin of Economics and Statistics* 72 (3):381.
Portnoy, S. 1991. Asymptotic behaviour of regression quantiles in non-Stationary, dependent cases. *Journal of Multivariate Analysis* 38 (1):100–13.
Powell, J. 1983. The asymptotic normality of two-Stage Least absolute deviations estimators. *Econometrica* 51 (5):1569–75.

Raheem, S. M. E., S. E. Ahmed, and K. A. Doksum. 2012. Absolute penalty and shrinkage estimation in partially linear models. *Computational Statistics & Data Analysis* 56 (4):874–91.

Ruppert, D., and R. J. Carroll. 1980. Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75 (372):828–38.

Sakata, S. 2007. Instrumental variable estimation based on conditional median restriction. *Journal of Econometrics* 141 (2):350–82.

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 (4):1580–92.

Stein, C. 1955. Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Berkeley, CA: University of California Press.

Tong, T., and Y. Wang. 2007. Optimal shrinkage estimation of variances with applications to microarray data analysis. *Journal of the American Statistical Association, March* 102 (477): 113–22. Vol. No.

White, H. 2001. *Asymptotic theory for econometricians*. Academic Press, San Diego.