1 Introduction

This paper considers estimating a structural equation using quantile regression with endogeneity problems. Since the seminal work by Koenker and Bassett (1978), the literature on quantile regressions has grown rapidly. There are two trends in the literature about quantile regression with endogeneity problems. The first one, denoted the 'structural approach', corresponds to models specified in terms of the conditional quantile of the structural equation. The second one, denoted the 'fitted-value approach', is based on the conditional quantile of the reduced-from equation. In this latter approach, the analysts substitute the endogenous regressors with their fitted values obtained from ancillary equations that are based on other exogenous variables.[1] In this paper, we show how to integrate these two approaches, while covering in particular the non-constant effect case fo quantile regressions.

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

The fitted-value approach, which we follow in this paper, is anchored on conditional quantile restrictions applied to the reduced-form equation. It leads to a simple two-step quantile regression analogous to the well-known 2SLS method. Such regression has been employed by empirical researchers, even though all theoretical results were not available[2]. Though, partial theoretical results had been explored by Amemiya (1982) and Powell (1983) who analyse the two-stage least-absolute-

[1] For 2SLS, an equivalent approach is the IV estimator based on the GMM principle.

[2] Arias et al. (2001), Garcia et al. (2001) and Chortareas et al. (2012).

deviations estimators in simple settings. Chen (1988) and Chen and Portnoy (1996) investigate two-stage quantile regressions in which the trimmed least squares (TLS) and LAD estimators are employed as the first-stage estimators, under an assumption of symmetric iid errors. TLS are found especially attractive because they are robust estimators (thanks to preliminary quantile regressions which trim outliers) that often preserve most of the efficiency of OLS. Kim and Muller (2004) use a similar approach with quantile regressions in the first stage. For linear equations estimated with quantile regressions under endogeneity, Chernozhukov and Hansen (2005) exploit the reduced-form for an inference procedure on structural conditional quantiles when the instruments are weak.

In this paper, we make several contributions. First, we derive the asymptotic distribution and the variance-covariance matrix of two-stage quantile estimators under very general conditions on both error terms and exogenous variables. These results were missing in the fitted-value approach literature. Second, we exhibit a 'bias transmission property' from the asymptotic representation of our estimator. We use this property to facilitate the analyses of the link of reduced-form and structural model, and to confine estimation bias on the intercept for some models. Third, we show how structural and fitted-values approaches can be integrated, providing some natural independence hypothesis for instruments. Fourth, we elicit the possibility of non-constant effects models with the fitted-value approach, a situation sometimes believed to be ruled out with this approach. Fifth, when the above independence hypothesis is met, the fitted-value approach allows the estimation of models that cannot be identified in the structural approach. We describe the role of relaxing the independence assumption on separating structural and fitted-value approaches. Sixth, we propose a new technique to improve the efficiency of two-stage quantile regressions.

This technique is based on an idea originally proposed by Amemiya (1982), where the composite dependent variable is a weighted combination of the original dependent variable and its fitted value.

A well-known method of reducing the variance of an estimator is to replace it by a weighted average of it with another estimator. In that case, the optimal weight is usually determined by minimising the (asymptotic) variance of the combined estimators[3]. This approach if often hard to apply to two-stage estimators because it generally requires the estimation of the joint distribution of the two estimators. In contrast, our approach of using instead a composite dependent variable yields an estimator such that: (i) consistency does not depend on the combination weight, and (ii) the asymptotic variance depends on the weight. In this setting, optimal weights can be obtained by minimising the asymptotic variance without perturbing the asymptotic properties of the estimator. Finally, Monte Carlo simulations are reported that exhibit a variety of small sample properties of our estimator, with and without variance reduction, and show considerable efficiency gains, in particular as compared to typical structural equation estimators.

The paper is organized as follows. Section 2 discusses the model and the assumptions. In Section 3, we derive the asymptotic representation of the two-stage quantile regression estimators. We characterise the asymptotic bias for general two-stage estimators in Section 4, where we also discuss the integration of structural and fitted-value approaches. We analyse in Section 5 the asymptotic normality and the asymptotic covariance matrix of two-stage quantile regressions based on LS or TLS predictions. We also investigate the optimal weights that minimize the asymptotic variance of the two-stage quantile regression estimators. In Section 6, we present Monte Carlo simulation results. Finally, Section 7 concludes. All the proofs are

[3] e.g., in James and Stein (1960), Sen and Saleh (1987), Kim and White (2001).

collected in Appendix A.

2 The Model

We are interested in the parameter ($\alpha_{0\theta}$) in the following equation for T observations and an arbitrary quantile index $\theta \in (0,1)$:

$$\begin{aligned} y_t &= x'_{1t}\beta_{0\theta} + Y'_t\gamma_{0\theta} + u_{t\theta} \\ &= z'_t\alpha_{0\theta} + u_{t\theta}, \end{aligned} \tag{1}$$

where $[y_t, Y'_t]$ is a $(G+1)$ row vector of endogenous variables, x'_{1t} is a K_1 row vector of exogenous variables, $z_t = [x'_{1t}, Y'_t]'$, $\alpha_{0\theta} = [\beta'_{0\theta}, \gamma'_{0\theta}]'$ and $u_{t\theta}$ is an error term. We emphasize that the coefficients $\beta_{0\theta}, \gamma_{0\theta}$ and the errors in this specification may vary with the considered quantile θ in order to allow for non-constant effects in quantile regressions introduced later on. We denote by x'_{2t} the row vector of the K_2 exogenous variables excluded from (1). We further assume that Y_t can be linearly predicted from the exogenous variables;

$$Y'_t = x'_t\Pi_0 + V'_t \tag{2}$$

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

By assumption, the first element of x_{1t} is 1 and the corresponding coefficient in $\beta_{0\theta}$ is non-zero[4]. This is a crucial assumption that will allow us in particular: (1) to specify non-constant effects under the fitted-value approach, and (2) to confine a bias to an intercept parameter, often less interesting for analysts than the slope coefficients. Besides, Jureckova (1984) shows that the absence of an intercept in the

[4]Using another coefficient of secondary interest is possible for our argument, while it is esier to present it by focusing on the intercept.

model would affect the asymptotic properties of quantile regressions, which suggests including an intercept. Using (1) and (2), y_t can also be expressed as:

$$y_t = x_t'\pi_{0\theta} + v_{t\theta}, \tag{3}$$

where

$$\begin{aligned}\pi_{0\theta} &= H(\Pi_0)\alpha_{0\theta} \text{ with } H(\Pi_0) = \left[\begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi_0\right], \\ v_{t\theta} &= u_{t\theta} + V_t'\gamma_{0\theta}.\end{aligned} \tag{4}$$

Moreover, we assume that the intercept coefficient in (3), denoted π_{00}, is non-zero. Equations (2) and (3) are the basis of the first-stage estimation that yields estimators $\hat{\pi}$ and $\hat{\Pi}$ respectively of π_0 and Π_0. So far, we did not mention any restriction on errors. The precise error restrictions will be introduced below in Assumptions 3, 4(ii) and 4$'$(ii) when dealing with examples of first-stage estimators. This is because we wish to keep the framework as general as possible until we deal with these examples. However, to set ideas, the reader may wish to consider conditional quantile restrictions on v_t in the fitted-value approach or on u_t in the structural approach.

A convenient way to generate randomness in a quantile regression model is to avail of the Skorohod representation of random variables, as in Chernozhukov and Hansen (2005). This would yield an equation of the type $y = q(x, U)$, where $U \sim \mathcal{U}_{(0,1)}$ and $q(x, \theta)$ denotes the θ-quantile conditional on x of variable y. Specialising this method to our setting, we obtain by analogy with eq. (1): $y_t = \beta_{00}(U) + \tilde{x}_{1t}'\beta_{01}(U) + Y_t'\gamma_0(U)$, in which the intercept term has been isolated and $\tilde{x}_{1t}'$ regroups all the variables of x_{1t}' except the intercept. The intercept term can itself been additively separated into a central tendency and a remainder: $\beta_{00}(U) = \bar{\beta}_{00} + \tilde{\beta}_{00}(U)$. Then, the component $\tilde{\beta}_{00}(U)$ can be considered as an additive error term in this model. The remaining (non intercept) parameters can be seen as random parameters since they depend on random U. Alternatively, they can be seen as part of the specification of non-constant effects for quantile regressions. A similar construction can be adopted for the reduced-form with $y_t = \bar{\pi}_{00} + \tilde{\pi}_{00}(U) + x_t'\pi_{01}(U)$, again with an intercept term and possibly non-constant effects. In such setting, $\tilde{\beta}_{00}(U)$ (respectively $\tilde{\pi}_{00}(U)$) is akin to our additive error term $u_{t\theta}$ (respectively $v_{t\theta}$). Of course, restrictions on $u_{t\theta}$, for example, would yield corresponding restrictions on U, through a change in variable and given specific distribution assumptions. The question of the correspondence between these respective specifications of the structural and the reduced-form models will be discussed in Section 4.

From now, we generally drop the dependence subscript of parameters and error terms with respect to θ, in order to alleviate notations. However, we keep in mind that these parameters and errors are still dependent on θ. We now specify the data generating process.

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

Studying quantile regressions with α-mixing processes is unusual. 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.

[5] The sequence $\{W_t\}$ of random variables is $\alpha-$mixing if $\alpha(s)$ decreases towards 0 as $s \to \infty$, where

$$\alpha(s) = \sup_t \sup_{A \in F_{-\infty}^t; B \in F_{t+s}^{\infty}} |P(A \cap B) - P(A)P(B)|$$

for $s \geq 1$ and F_s^t denote the σ-field generated by $(W_s, \ldots, W_t)$ for $-\infty \leq s \leq t \leq \infty$. The sequence is called $\alpha-$mixing of size $-a$ if $\alpha(s) = O(s^{-a-\varepsilon})$ for some $\varepsilon > 0$.

It is generally possible to employ unbiased estimators in the first stage. However, in order to exploit later on a trade-off between bias and efficiency, we allow in Assumption 2 for inconsistent first-stage estimation with bounded bias terms. It is done in a form convenient for including the contribution of first-stage estimators in the asymptotic distribution of our final estimator, instead of just specifying the stochastic limits of $\hat{\pi}$ and $\hat{\Pi}$. The precise restrictions on v_t and V_t corresponding to π_0 and Π_0 will be brought up later on.

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

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

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

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

where $\hat{y}_t = x_t'\hat{\pi}$ and q is a positive scalar constant. In the quantile regression in (5), the dependent variable $q y_t + (1-q)\hat{y}_t$ is a weighted average of y_t and its fitted-value

$\hat{y}_t$ obtained from the reduced form equation in (3). The combination weight q is restricted to be positive for a technical reason discussed in the proof of Proposition 1 below. Alternatively, as in Powell (1983), the case q negative is also possible by imposing $\theta = 0.5$, i.e. with the LAD estimator. The reformulation of the dependent variable as $qy_t + (1-q)\hat{y}_t$ was originally suggested by Amemiya (1982) to improve efficiency in two-stage estimation with $0 < q < 1$. The case $q = 1$ corresponds to the usual two-stage quantile regression estimator, while $q = 0$ corresponds to the inverse regression estimator under exact identification. Thus, the new dependent variable introduces a trade-off between two estimation methods. Our analysis is based on the asymptotic representation of the 2SQR(θ, q) under the following sufficient conditions.

Assumption 3 (i) *$H(\Pi_0 + B_\Pi)$ is of full column rank.*
(ii) *Let $F_t(.|x)$ be the conditional cdf and $f_t(.|x)$ be the conditional 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 ($f_t(\cdot|x) < f_0$, for all x).*
(iii) *The matrices $Q = \lim\limits_{T\to\infty} E\left[\frac{1}{T}\sum\limits_{t=1}^{T} x_t x_t'\right]$ and $Q_0 = \lim\limits_{T\to\infty} E\left[\frac{1}{T}\sum\limits_{t=1}^{T} f_t(0|x_t) x_t x_t'\right]$ are finite and positive definite.*
(iv) *$E(\psi_\theta(v_t)|x_t) = 0$, for any arbitrary θ.*
(v) *$\exists\, C > 0, \forall\, t, E(\|x_t\|^3) < C < \infty$.*

Assumption 3(i) is analogous to the usual identification condition for simultaneous equations models. The bias B_Π appears in the condition because the first-stage estimator converges towards $\Pi_0 + B_\Pi$. In the case when OLS is used for estimating Π_0, Assumption 3(i) ensures that $E\left[x_t Y_t\right] \neq 0$. It also implies similar conditions when other estimators are used. Assumption 3(ii) simplifies the demonstration of convergence of remainder terms to zero for the calculation of the asymptotic representation.

The second part of Assumption 3(iii) is the counterpart of the usual condition for OLS that the sample second moment matrix of the regressor vectors converges towards a finite positive definite matrix, which corresponds to the first part. The last condition is akin to the one in the conventional IV approach in that this condition is necessary for consistency and for the inversion of the relevant empirical process to establish the asymptotic normality.

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

Assumption 3(iv) is central to our fitted-value approach in which the conditional quantile restriction is placed on the reduced-form error v_t and the information set used for the conditional restriction exclusively consists of exogenous variables x_t. It has been used in simpler settings in Amemiya (1982), Powell (1981), Chen and Portnoy (1998) and Kim and Muller (2004). On the other hand, a typical conditional quantile restriction employed in the structural approach would be that the conditional quantile restriction is on the structural error term u_t and the information set includes the set of endogenous variables Y_t. Although some researchers may find it more intuitive to put the conditional quantile restriction directly on the structural error term, the fitted value approach provides an alternative and convenient way of exploring the

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

conditional distribution of the dependent variable y_t through the reduced-form error: that is, indirectly. The structural effect, which is the main interest, is then recovered through the second stage of the estimation. In that sense, the fitted-value approach is somewhat akin to that of the indirect inference literature where inference about a model raising estimation difficulties is done through its link with another model easier to handle.

The fitted-value approach has several advantages on approaches based on conditional quantiles of u_t. First, the 2SQR can be characterised by an explicit asymptotic representation that is liable to tractable analyses, as we shall demonstrate. This alone much facilitates the understanding of the properties of the estimator. Second, methods based on conditional quantiles of u_t may lead to computational complications such as: simulation techniques (Chernozhukov and Hansen, 2006), preliminary nonparametric estimation (Abadie et al., 2002, Chen et al., 2003, Lee, 2007), grid search (Chernozhukov and Hansen, 2005). As a matter of fact, it seems currently impossible to use these methods with large data sets when more than very few conditioning variables occur in the model, due to excessive computation burden. This is not the case with the straightforward fitted-value approach. Third, and this is a major point of this paper, considering 2SQR allows us to discuss a new powerful method of variance reduction.

However, the fitted-value approach has a major shortcoming: It does not seem to allow intuitive interpretation of the implied restrictions in the structural model. In particular, considering together structural and reduced-form equations raises the question of the coordinating the specifications of the respective quantile regressions. We shall show that, provided an independence assumption, the quantile restrictions on structural and reduced-form equations are actually equivalent, as soon as a proper definition of error terms is adopted. We now study the asymptotic properties of the 2SQR(θ, q).

3 The Asymptotic Representation

To derive the asymptotic representation of the 2SQR(θ, q), we define an empirical process given by

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

where Δ is a $K \times 1$ vector.

Applying Theorem II.8 in Andrews (1990) yields the following lemma. The lemma is proven only for the quantile regression case, while similar derivations can be done for other two-stage M-estimators.

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

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

Combining Lemma 1 and Assumption 2 allows us to obtain the asymptotic representation for the 2SQR(θ, q) with a possible bias term B_α. The consistency of $\hat{\alpha}$ to $\alpha_0 + B_\alpha$ is a by-product of the following proposition.

Proposition 1 *Under Assumptions 1-3,*

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

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

The asymptotic representation of the 2SQR(θ, q) is composed of four additive right-hand-side terms[7]. 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 the uncertainty of the estimator. Then, if $\hat{\pi}$ and $\hat{\Pi}$ are consistent, it is straightforward to show that the 2SQR(θ, q) is consistent. If $q = 1$, the influence of $\hat{\pi}$ vanishes. The presence of the contribution of $\hat{\pi}$ may imply contradictions between some chosen restrictions on errors in the first and second stages and cause biases, which will be explained in detail in the next section. The formula of B_α is obtained as the value allowing $T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1)$, and is derived from the first-order conditions of the second-stage estimation.

4 The Asymptotic Bias

We now show that asymptotic biases on the coefficients of the first-stage estimators corresponding to the exogenous variables in the structural equation are transmitted to the estimated two-stage coefficients of the same variables exclusively and integrally. This is useful because, as we shall show, interesting cases exist that correspond to first-stage biases only on the intercept. In particular, we shall exploit this property to handle a trade-off between efficiency and an often little damaging bias. The sufficient stochastic assumptions to obtain this characterisation of the bias transmission are very general, including general serial correlations and heteroskedasticity.

[7] Other derivations of asymptotic representations of quantile regression estimators have been developed (Phillips, 1991, Pollard, 1991), which involve slightly different assumptions. Other possible approaches to the asymptotic representation of 2SQR are in Chen et al. (2003) and Chernozhukov and Hansen (2005).

Proposition 2 *Assume that the bias B_π and B_Π are restricted to be possibly non-zero only for their first K_1 components, then we have* $B_\alpha = \begin{bmatrix} (1-q)B\pi - B_\Pi\gamma_0 \\ 0_G \end{bmatrix}$,

where 0_G is a G vector of zeros.

Situations where the asymptotic bias of the first-stage estimators exclusively affects the intercept term are interesting in that empirical researchers may often pay little attention to this term and rather base their analyses on the estimates of the slope coefficients that carry more explanatory meaning. Then, we focus on the case where the asymptotic bias is only present in the first-stage intercept estimator. In that case, Proposition 2 implies that the only coordinate of $\hat{\alpha}$ with a possible asymptotic bias is the intercept. Moreover, this asymptotic bias is equal to $(1-q)$ times the asymptotic bias in the intercept in $\hat{\pi}$ minus the asymptotic bias in the intercept in $\hat{\Pi}\gamma_0$. In these conditions, there is no difficulty in achieving the consistency and asymptotic normality for the slope estimators of interest using Proposition 1.

In situations where $\hat{\Pi}$ is not asymptotically biased, choosing $q = 1$ guarantees that there is no bias. The first-stage estimation methods can be chosen to eliminate the biases on $\hat{\pi}$ and $\hat{\Pi}$ (e.g., by using the same quantile regressions in the two stages as in Kim and Muller, 2004). However, the researcher may also choose the first-stage estimation methods for her own reasons, for example because there already exists some available estimation results. In this paper, we propose choosing $q \neq 1$ and selecting first-stage estimators so as to improve the final slope estimator efficiency, while allowing for an asymptotic bias on the intercept.

Proposition 1 can also be used to generate some valid insight about the relationship between structural and reduced-form models. Let us start again with equations (1) with possible non-constant effects and (2) under the conditional quantile restriction $E(\psi_\theta(u_{t\theta})|x_t) = 0$, akin to the structural quantile restriction in Chernozhukov and

Hansen (2005) and $E(V_t|x_t) = 0$ to set ideas with OLS estimates of the prediction equation. Then, the reduced-form equation is (3) with (4) holding. The corresponding restriction for $v_{t\theta}$ is $E(\psi_\theta(v_{t\theta} - V_t'\gamma_0)|x_t) = 0$, little liable to interpretation, and this does not yield any obvious characterisation of $E(\psi_\theta(v_{t\theta})|x_t)$.

However, we can separate some centred intercept term in the reduced-form by denoting

$y_t = x_t'\pi_{0\theta} + v_{t\theta} = \pi_{00} + F^{-1}_{v_{t\theta}|x_t}(\theta) + \tilde{x}_t'\pi_{01\theta} + v^*_{t\theta}$, where $\tilde{x}_t$ regroups all non-constant variables in x_t, with the corresponding parameter vector denoted $\pi_{01\theta}$, π_{00} is a fixed-intercept parameter not depending on θ, and $v^*_{t\theta} = v_{t\theta} - F^{-1}_{v_{t\theta}|x_t}(\theta)$. Let us now look at the conditional quantile restriction characterising $v^*_{t\theta}$. We have $E(\psi_\theta(v^*_{t\theta})|x_t) = \theta - P\left[v^*_{t\theta} \leq 0|x_t\right] = \theta - P\left[v_{t\theta} \leq F^{-1}_{v_{t\theta}|x_t}(\theta)|x_t\right] = \theta - \theta = 0$. As a consequence, we obtain the reduced-form quantile regression restriction, provided we accept the introduction of a possible nuisance bias term $F^{-1}_{v_{t\theta}|x_t}(\theta)$ that may affect all coefficients of the model when it is linear in x_t, or even be nonlinear in x_t.

Let us now assume that $u_{t\theta}$ and V_t are independent of $\tilde{x}_t$, that is, of x_t except constant variables. Although such characterisation of instrumental variables may be deemed to be strong by some authors, it is usually the way instrumental variables are intuitively found by empiricists: variables that are not connected at all with the model errors seen as a remainder of the explanation of the dependent variable given the effects of explanatory variables. Weaker orthogonality conditions, for example used for 2SLS, are often just a consequence of such intuitive selection of plausible instruments. Note that U independent of x_t in Chernozhukov and Hansen's setting would naturally imply $u_{t\theta}$ independent of $\tilde{x}_t$ in our setting since $u_{t\theta}$ can be seen as a simple transformation of U.

Under this assumption of independence, we have $F^{-1}_{v_{t\theta}|x_t} = F^{-1}_{v_{t\theta}}$ and the perturbation caused by this term is confined to the intercept. We obtain: $y_t = \pi_{00} + F^{-1}_{v_{t\theta}}(\theta) + \tilde{x}_t'\pi_{01\theta} + v^*_{t\theta}$. What is remarkable here is that the shift in the reduced-form parameters, when estimating a reduced-form quantile regression based on $E(\psi_\theta(v^*_{t\theta})|x_t) = 0$ instead of $E(\psi_\theta(v_{t\theta})|x_t) = 0$, is confined to the intercept to which the difference between $v^*_{t\theta}$ and $v_{t\theta}$ amounts. Then, according to Proposition 2, a bias is generated exclusively on the intercept term of the structural model.

Reciprocally, it is easy to see that starting from $E(\psi_\theta(v_{t\theta})|x_t) = 0$ and assuming the independence of v_t and V_t with respect to $\tilde{x}_t$, we can obtain the structural restriction $E(\psi_\theta(u_{t\theta})|x_t) = 0$ for a structural model with the right value of parameters, except perhaps for a bias on the intercept.

However, the above independence condition also implies that $v_{t\theta}$ is independent of $\tilde{x}_t$. This independence delivers a reduced-form quantile regression that is restricted to have constant effects, that is with the same coefficients (except the intercept) for any quantile θ. Because of the linear relationship between structural and reduced-form parameters that is embedded in (4), the structural quantile regression is also characterised by constant-effects. As a matter of fact, there is a tension between the need of linking structural and reduced-form models and the wish of having non-constant effects.

We now deal with this matter by relaxing the link between models through allowing possible biases on some other coefficients. Specifically, we now only assume that $F^{-1}_{v_{t\theta}|x_t}(\theta) = F^{-1}_{v_{t\theta}|x_{1t}}(\theta)$, which is satisfied for any θ as long as $v_{t\theta}$ is independent on x_{2t}. This latter condition implies that there are constant effects in the reduced-form equation for the coefficients of the x_{2t}, but not necessarily for the coefficients of the x_{1t}. It is also equivalent with $u_{t\theta}$ and V_t independent on x_{2t}. It is appropriate that it corresponds to a natural 'instrumental variable' condition for the structural model and x_{2t}. In this setting, the bias can now be confined to the first K_1 variables in the reduced-form model instead of the intercept only. Indeed,

$y_t = x_{1t}'\pi_{0K_1\theta} + x_{2t}'\pi_{0K_2\theta} + v_{t\theta} = x_{1t}'\pi_{0K_1\theta} + x_{2t}'\pi_{0K_2\theta} + F^{-1}_{v_{t\theta}|x_{1t}}(\theta) + v^*_{t\theta} = x_{1t}'\pi^*_{0K_1\theta} + x_{2t}'\pi_{0K_2\theta} + v^*_{t\theta}$,

where $\pi_{0K_1\theta}$ and $\pi_{0K_2\theta}$ denote the parameter vectors respectively associated with x_{1t} and x_{2t} in the reduced-form, $v^*_{t\theta} = v_{t\theta} + F^{-1}_{v_{t\theta}|x_{1t}}$, $\pi^*_{0K_1\theta}$ denotes the unbiased parmeter vector associated with x_{1t}. Then, a routine quantile regression estimation of the reduced-form, based on $E(\psi_\theta(v_{t\theta})|x_t) = 0$, would yield unbiased estimates for the K_2 last coefficients, as it is explicit by rewriting the restriction as $E(\psi_\theta(y_t - x_{1t}'\pi^*_{0K_1\theta} - x_{2t}'\pi_{0K_2\theta} - F^{-1}_{v_{t\theta}|x_{1t}}(\theta))|x_t) = 0$.

Allowing a bias on the parameters $\pi_{0K_1\theta}$ enables us to introduce non-constant effects on the vector $\beta_{0\theta}$, even though these parameters are biased due to the bias transmission result. This alone would be a generalisation of the *stricto sensu* constant effect structural quantile regression, which may be useful if the researcher's interest is concentrated on vector γ_0 that can be estimated consistently. However, solving the system of linear equations between parameters in (4) implies that there is in general an influence of the estimation of $\pi_{0K_1\theta}$ on that of γ_0, and this without involving the bias on $\pi_{0K_1\theta}$, due to Proposition 2. Therefore, specifying non-constant effects for $\pi^*_{0K_1\theta}$ (e.g., with a given functional form of θ) implies in general non-constant effects for γ_0. As a matter of fact, only allowing for biases on the intercept plus on another component of $\pi_{0K_1\theta}$ would be enough to obtain this result. There is therefore a variety of independence conditions that could be used to generate non-constant structural effects without giving up too many consistency results for the estimation of the exogenous variables parameters in the structural model. Note that the same approach is not possible by allowing only a bias on the intercept as in our setting the original intercept has been used to generate the additive error term $v_{t\theta}$ and the remaining fixed intercept term cannot vary with θ.

These results have considerable consequences for the interpretation of the fitted-value approach. Namely, under the independence hypothesis, the estimates based on the conditional quantile of the reduced-form can be used to recover at least the slope estimates of the conditional quantile of the structural form. Assumption A3(iv) can therefore, under these conditions, be considered as a useful characterisation of the structural conditional quantile distribution.

There is more. Referring to Proposition 2, providing the independence assumption, any property of the slope coefficients that is invariant to linear transformations is preserved when moving from the reduced-form model to the structural form model and vice versa. In particular, constant (respectively, non-constant) effects specifications are equivalent for both models. That is: if there are constant (respectively, non-constant) effects for the reduced-form, they are also constant (non-constant) for the structural form, and vice versa, all this under the independence condition.

A related question is: what is the weakest 'independence' condition to impose in order to obtain these convenient properties of specification coordination for structural and reduced-form models. As above we have $E(\psi_\theta(v^*_{t\theta})|x_t) = 0$, even without any independence condition. However, without some kind of independence condition, parameter π_0 cannot be recovered because what is estimated is $x_t'\pi_{0\theta} + F^{-1}_{v_{t\theta}|x_t}(\theta)$ and the contribution of the two additive terms cannot be distinguished. In contrast, researchers could accept that $\pi_{00\theta}$ could not be recovered as it is after all only a little interesting intercept. That is what we have under the independence assumption using the bias transmission property, since what is estimated is $\pi_{00\theta} + F^{-1}_{v_{t\theta}}(\theta)$ instead of the intercept.

Then, the minimal condition to identify the slope coefficients $\pi_{10\theta}$ is: $F^{-1}_{v_{t\theta}|\tilde{x}_t}(\theta)$ is known (or can be consistently estimated). This condition is for example satisfied when $v_{t\theta}$ is independent of $\tilde{x}_t$. Another favourable possibility is $F^{-1}_{v_{t\theta}|\tilde{x}_t}(\theta) = F^{-1}_{v_{t\theta}}(\theta) + x_t'\delta$, where δ is known.

In all these cases, the slope coefficients in $\pi_{01\theta}$ are estimated consistently in the reduced-form quantile regression. Then, the bias transmission property yields consistent estimators of α_0, except perhaps for the intercept. Similar properties can be derived by confining the bias to the first K_1 components of x_t instead.

Finally, it may be interesting to deal with cases where no useful weak or strong independence hypothesis is satisfied. In that cases, the reduced-form quantile restrictions may not correspond to any structural quantile restriction. However, even if their interpretation in terms of the structural model is less obvious, they may still be interesting to explore. In particular, they may correspond to situations where no condition for identifying the structural quantile model are known (e.g., Chernozhukov and Hansen independence condition is not satisfied either). In these situations, investigating the reduced-form model may still remain of interest, as well as examining the corresponding pseudo-estimates for the structural model.

Let us now compare more precisely Chernozhukov and Hansen (2005) approach and ours. An identification restriction of a structural quantile regression under endogeneity of Y_t is $E(\psi_\theta(u_{t\theta})|x_t) = 0$. This condition can be rewritten as

$P\left[y_t \leq \bar{\beta}_{00} + \tilde{x}_{1t}'\beta_{10}(\theta) + Y_t'\gamma(\theta)|x_t\right] = \theta$. As a comparison benchmark, the condition by Chernozhukov and Hansen with our notations would be $P\left[y_t \leq \beta_{00}(\theta) + \tilde{x}_{1t}'\beta_{10}(\theta) + Y_t'\gamma(\theta\right.$ θ, where $\beta_{00}(\theta) = \bar{\beta}_{00} + u_t$. The respective terms at the right-hand side of the inequality in the probability bracket represent two distinct models of quantiles of y_t

conditional on x_{1t} and Y_t. Linking with Chernozhukov and Hansen's notations, we have $u_t = F^{-1}_{U_t|x_t}(\theta)$.

The main difference between these two models is that we do not allow for models without intercept or models with intercept terms varying with quantiles. We also employ an explicit additive error term u_t, extracted from the intercept, so as to imposes restrictions on this error term only and leave the random parameters $\beta_{10}(U)$ and $\gamma(U)$ a priori unrestricted with respect to instruments. However, imposing our above independence assumption may be used to complete the restrictions and implying the independence of both $\beta_{10}(U)$ and $\gamma(U)$ with respect to x_t. The latter assumption allows for some coordination of the quantile restrictions on u_t with the specification of random parameters $\beta_{10}(U)$ and $\gamma(U)$.

We make one further step in distinguing us from Chernozhukov and Hansen. Namely, under endogeneity we consider $P\left[y_t \leq \bar{\pi}_{00} + x_t' \pi_{10}(\theta) \mid x_t\right] = \theta$ instead of their condition. In that case, we have with obvious notations: $\pi_{0\theta} = H(\Pi_0)\alpha_{0\theta}$ with

$$\begin{bmatrix} \bar{\pi}_{00} \\ \pi_{110}(\theta) \\ \pi_{120}(\theta) \end{bmatrix} = \left[\begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi_0 \right] \begin{bmatrix} \bar{\beta}_{00} \\ \beta_{10}(\theta) \\ \gamma(\theta) \end{bmatrix}.$$

Then, we obtain $P\left[y_t \leq \bar{\beta}_{00} + \tilde{x}_{1t}' \beta_{10}(\theta) + x_t' \Pi_0 \gamma(\theta) \mid x_t\right] = \theta$, which expresses the content of the fitted-value method since Y_t has been replaced by its true fitted value. This restriction corresponds to what we estimate, and allows the identification of the structural slope parameters as we explained before. Our restrictions are first imposed on v_t and V_t only, which then generates coordinated restrictions on random parameters $\beta_{10}(\theta)$ and $\gamma(\theta)$ through the assumed independence of both v_t and V_t with respect to x_t and the bias transmission property. We now turn to the asymptotic covariance matrix for 2SQR(θ, q) with some specific first-stage estimators.

5 Asymptotic Normality and Covariance Matrix with LS and Trimmed-LS Predictions

In this section, we examine the use of (non-robust) LS estimation and (robust) Trimmed-Least-Squares (TLS) estimation of π_0 and Π_0 in the first step of 2SQR(θ, q). There are several reasons to consider LS estimators. First, LS estimators are popular and available LS estimation results for the first-stage equations may be ready to be used. Second, the researcher may wish to use quantile regressions not for their robustness but rather for focusing on a given location of the conditional distribution of the dependent variable. Then, using LS estimators as a first stage may improve the efficiency of the estimation procedure. Finally, that is an approach that empirical researchers have been using in practice[8]. Alternatively, using TLS in the first stage guarantees the robustness of this estimation stage, while some efficiency may be lost. Using twice the same quantile regression in both stages has been examined in Kim and Muller (2004). In that case, there is no consistency issue, but no opportunity for asymptotic variance reduction either.

An issue here is that the bias term $E(v_t)$ is not necessarily zero because it conflicts with the restriction $E(\psi_\theta(v_t)) = 0$, which is implied by Assumption 3(iv); that is, the θ^{th} quantile and the mean of v_t cannot be zero at the same time. To be able to use the usual Bahadur representation of OLS on the intercept, we define the centered errors $v_t^* = v_t - E(v_t|x_t)$, $u_t^* = u_t - E(u_t|x_t)$ and $V_t^* = V_t - E(V_t|x_t)$. By construction, $E(v_t^*|x_t) = E(u_t^*|x_t) = E(V_t^*|x_t) = 0$. Moreover, we have $u_t^* = v_t^* - V_t^{*\prime}\gamma_0$. Restrictions on univariate errors u_t^* and v_t^* are sufficient, instead of restrictions on u_t^* and all the components of V_t^*. We impose the following orthogonality conditions on v_t and V_t.

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

This is more restrictive than what is necessary for Assumption 2 or for Proposition 2, because we now want to confine the possible bias to the intercept exclusively.

Assumption 3$'$ (i) $E(v_t|x_{(j)t}) = E(v_t), j = 2, ..., K,$ where $x_{(j)t}$ denotes the j^{th} component of x_t.
(ii) $E(u_t|x_{(j)t}) = E(u_t), j = 2, ..., K.$

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

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

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

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

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

Assumption 3$'$(i) (respectively (ii)) imposes the orthogonality of the reduced-form (respectively structural) errors with all non-constant exogenous variables. It slightly strengthens the exogeneity requirement in Assumption 3(iv) so that the bias can be confined to the intercept.

These assumptions impose only light restrictions on V_t when they are several endogenous variables. Indeed, our results for the estimates of the structural equation will be valid under any restrictions on V_t compatible with the restrictions that are imposed on u_t or v_t. The bias B_π is generally non-zero for $q \neq 1$. In contrast, B_Π can be non-zero or not, even with $q = 1$, depending on the restrictions imposed on V_t. In the case $q = 1$, a natural specification suggests $E(V_t|x_t) = 0$ while using OLS to estimate (2) and no bias at all. In other cases, B_π and B_Π may have to be taken into account.

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

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

Owing to the characterisation of B_π and B_Π and Propositions 1 and 2, we have $B_\alpha = ((1-q)E(v_t) - E(V_t')\gamma_0, 0, \ldots, 0)'$. The intercept estimator may be asymptotically biased, while the slope estimators are not. Meanwhile, the asymptotic normality of $\tilde{\alpha} - \alpha_0 - B_\alpha$ can be derived under the following assumptions.

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

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

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

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

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

These asymptotic properties of the 2SQR(θ, q) have been established for a given value of q. To improve efficiency, q can be replaced with its optimal value (q^*) obtained by minimising the asymptotic covariance matrix that is shown in Proposition 3. However, there are many ways of minimising a multi-dimensional covariance matrix. For example, one may wish to minimise some norm of the matrix (e.g. the mean square error). One may also wish to minimise the standard error for a given coefficient of interest in the structural model. For all these procedures, in the iid case where the effect of q is concentrated in a scalar function, a unique and explicit solution q^* can be obtained. In the general case, q^* can also be made explicit when the MSE is minimised. Consistent preliminary estimators of q^* do not perturb the asymptotic properties of the 2SQR, which can be characterised as a MINPIN estimator (Andrews, 1994, p. 2263), as long as a stochastic equicontinuity condition of the global empirical process is valid.

In the case of least-squares plus quantile regression estimation, choosing values for q different from 1 introduces an asymptotic bias confined to the intercept term, while it allows the values for q to be selected so as to reduce the asymptotic variance of the consistent slope estimators. Since the bias can be easily corrected as its formula is known, our approach may improve efficiency in two-stage estimation.

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

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

A consistent estimator for q^* is obtained by substituting a consistent kernel-estimator

$\hat{f}(0)$ for $f(0)$, and residuals for error terms;

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

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

To address robustness concerns, we now propose an estimator based on a robust first-stage estimator: the symmetrically trimmed-LS estimator (TLS). The TLS of π in the model $y = X\pi + v$ is $\hat{\pi}_{TLS} = (X'AX)^{-}XAy$, where $A = (a_{ij}), i, j = 1, ..., p$ and $a_{ij} = I_{\left[i=j \text{ and } X_i'\hat{\pi}(\mu) < y_i < X_i'\hat{\pi}(1-\mu)\right]}$, $\hat{\pi}(\mu)$ is the quantile regression estimator centered on a given quantile μ to be chosen a priori. Chen and Portnoy (1996) provide the TLS Bahadur representation. Let $\breve{\alpha}$ 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) $E(v_t|x_{(j)t}) - \mu\left[F^{-1}_{v|x_{(j)t}}(\mu) + F^{-1}_{v|x_{(j)t}}(1-\mu)\right] = 0.$
(ii) $E(u_t|x_{(j)t}) - \mu\left[F^{-1}_{u|x_{(j)t}}(\mu) + F^{-1}_{u|x_{(j)t}}(1-\mu)\right] = 0.$
Assumption $4'$ (i) *There exist finite constants* $\Delta_{\tilde{u}}$ *and* $\Delta_{\tilde{v}}$ *such that* $E|x_{ti}\tilde{u}_t^*|^3 < \Delta_{\tilde{u}}$ *and* $E|x_{ti}\tilde{v}_t^*|^3 < \Delta_{\tilde{v}}$ *, for all i and t, where* $\tilde{v}_t^* = v_t - \mu\left[F^{-1}_{v|x_t}(\mu) + F^{-1}_{v|x_t}(1-\mu)\right], \tilde{V}_{jt}^* = V_{jt} - \mu\left[F^{-1}_{V_j|x_t}(\mu) + F^{-1}_{V_j|x_t}(1-\mu)\right].$
(ii) The covariance matrix $\tilde{V}_T = var\left(T^{-1/2}\sum_{t=1}^{T}\tilde{S}_t\right)$ *is positive definite for T sufficiently large, where* $\tilde{S}_t = (q\psi_\theta(v_t), q\tilde{v}_t^* - \tilde{u}_t^*)' \otimes x_t$.

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

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

where $\tilde{B}_\alpha = \left[(1-q)\tilde{B}_\pi - \tilde{B}_\Pi\gamma\right]'$, $\tilde{B}_\pi = (E(v_t) - \mu\left[F_v^{-1}(\mu) + F_v^{-1}(1-\mu)\right], 0, \ldots, 0)'$, $\tilde{B}_{\Pi_j} = (E(V_{jt}) - \mu\left[F_{V_j}^{-1}(\mu) + F_{V_j}^{-1}(1-\mu)\right], 0, \ldots, 0)'$, $\tilde{B}_\Pi = \left(\tilde{B}_{\Pi_j}\right), j = 1, ..., K_1$. The asymptotic representation together with Assumptions $3''$ and $4'$ delivers the following asymptotic normality of $\breve{\alpha}$.

Proposition 4 *Under Assumptions 1,3, $3'$ and $4'$:*

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

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

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

$$
q^* = \frac{E(\tilde{v}_t^*\tilde{u}_t^*) - f(0)^{-1}E(\psi_\theta(v_t)\tilde{u}_t^*)}{f(0)^{-2}\theta(1-\theta) + E(\tilde{v}_t^{*2}) - 2f(0)^{-1}E(\psi_\theta(v_t)\tilde{v}_t^*)}. \tag{10}
$$

A consistent estimator for q^* is

$$
\hat{q} = \frac{\sum_{t=1}^{T}\breve{v}_t^*\breve{u}_t^* - \hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\breve{u}_t^*}{T\hat{f}(0)^{-2}\theta(1-\theta) + \sum_{t=1}^{T}\breve{v}_t^{*2} - 2\hat{f}(0)^{-1}\sum_{t=1}^{T}\psi_\theta(\hat{v}_t)\breve{v}_t^*} \tag{11}
$$

where $\breve{u}_t^* = \breve{v}_t^* - \breve{V}_t^{*\prime}\breve{\gamma}$, $\breve{v}_t^* = y_t - x_t'\hat{\pi}_{TLS}$, $\breve{V}_t^{*\prime} = Y_t' - x_t'\hat{\Pi}_{TLS}$, $\hat{v}_t = y_t - x_t'\hat{\pi}_\theta$ and $\hat{\pi}_\theta = \arg\min_{\pi} \sum_{t=1}^{T}\rho_\theta(y_t - x_t'\pi)$. In the next section, we present Monte Carlo simulation results showing how much variance reduction can be realized in finite samples with our method.[9]

6 Monte Carlo Simulations

6.1 Simulation Set-up

The data generating process used in the simulations is described in Appendix B. We study the finite sample properties of our two proposed two-stage estimators: (1) the OLS plus quantile regression estimator (2SQR1), and (2) the TLS plus quantile regression estimator (2SQR2). We impose $E\left(\psi_\theta(v_t)|x_t\right) = 0$ for a given θ. That is: for each θ, we re-generate the error terms such that $E\left(\psi_\theta(v_t)|x_t\right) = 0$ is satisfied, which means that we consider different models associated with the different chosen quantiles θ for Assumption 3(iv). The equation of interest is assumed to be over-identified and the parameter values are set to $\beta' = (\beta_{0,1}, \beta_{0,2}) = (1, 0.2)$ and $\gamma = 0.5$. We generate the error terms by using three alternative distributions: the standard normal N(0,1), the Student-t with 3 degrees of freedom $t(3)$ and the Lognormal LN(0,1). The exogenous variables x_t are drawn independently from the errors from a normal distribution. The total number of replications is 1,000. For each replication, 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. In the iid case, the optimal value q^* is obtained by simulating the formula in (8) or (10), while $\hat{q}$ is estimated through (9) or (11).

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

6.2 Results

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

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. Here, 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 $q = \hat{q}$; an impressive efficiency gain. However, there is virtually no efficiency gain with small values of θ less than around 0.5.

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

Given the generally substantial efficiency gains, it is natural to ask how close the reduced variance is to the Cramer-Rao lower bound. We have calculated the CR bound numerically for each distribution in Tables 4-5. Table 4 shows the simulated asymptotic standard deviations for 2SLS, 2SQR1$(\theta, \hat{q})$ with $\theta = 0.25, 0.50, 0.95$, along with the simulated CR bounds. We only discuss the slope coefficients as the intercept coefficient may be biased. In the normal case, the 2SLS is very close to the simulated CR bound for $T = 300$. With $T = 50$, the 2SLS efficiency loss is no longer negligible. The 2SQR1$(\theta, \hat{q})$ is almost as accurate as the 2SLS for $T = 300$, for any considered quantile. Moreover, for $T = 50$, the 2SQR1$(\theta, \hat{q})$ remains close to the CR bound. This is a nicely surprising result given the usual inefficiency of quantile regressions. Here, reformulating the dependent variable allows us to reach almost optimal efficiency, while applying a typically inefficient second-stage estimator.

With Student errors, the 2SLS yields large efficiency loss. For all T and $\theta = 0.5$ and 0.25, using our method instead yields large accuracy gain, while still relatively far from the bound, especially for $\theta = 0.95$. Finally, under lognormality, 2SLS performance is always poor. This is less so for the 2SQR1$(\theta, \hat{q})$. Here again, choosing $\theta = 0.25$ considerably improves efficiency. This is not surprising since asymmetric distributions are typically difficult to deal with. A logarithmic transformation of the dependent variable could be an alternative strategy in this situation. When this is not possible, our method alleviates the efficiency loss for quantiles close to the mode of the error distribution and for central quantiles.

Let us now turn to 2SQR2 based on the TLS at the first stage (results in Tables 4(b) and 5(b)). We have also tried the LAD-2SQR(θ) (i.e. LAD in the first-stage as in Chen and Portnoy, 1996). However, the results are almost identical to that of the 2SQR2 to which we limit our comments. What seems to matter here is the robustness of the first stage more than the chosen estimator. Trimming the 2SQR2 at quantile $\varphi = 0.25$ yields more accurate results than trimming at 0.05 or 0.10[11] and for $T = 300$, trimming at $0.05, 0.10$ or 0.25 is almost indifferent. Hence, we focus on the case $\varphi = 0.25$.

While our above improved estimator 2SQR1$(\theta, \hat{q})$ generally performs better than the previously considered alternative estimators, it is often outperformed by the estimators introduced by Chen and Portnoy (2SQR2 with $q = 1$). However, these estimators can be further improved by redefining the dependent variable as we propose. The obtained estimator 2SQR2$(\theta, \hat{q})$ dominates or is almost equivalent to all the considered competitors in the studied cases.

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

Under normal errors, the 2SQR1$(\theta, \hat{q})$ and the 2SQR2$(\theta, \hat{q})$ both almost reach the CR bound, whatever the considered quantile. Here, reformulating the dependent variable is fruitful, especially for upper quantiles for which it allows massive efficiency gains. The 2SQR2$(\theta, q = 1)$ is slightly outperformed by the 2SQR1$(\theta, \hat{q})$, perhaps because trimming here only discards information. With Student errors, the 2SQR2$(\theta, \hat{q})$ is often the more accurate estimator, yielding results fairly close to the bound in the constant-effects case. Nevertheless, the 2SQR2$(\theta, 1)$ remains an equivalent solution for middle quantiles.

[11] Except for normal errors and $T = 50$ in which case trimming at 0.25 is only slightly inferior.

Under lognormality, no studied estimator approaches the CR bound in the constant-effect case. However, using the 2SQR2$(\theta, \hat{q})$ generally yields the best accuracy. For upper quantiles, reformulating the dependent variables delivers huge efficiency gains.

It is interesting to reflect on the proximity of the results of the 2SQR1$(\theta, \hat{q})$ and the 2SQR2$(\theta, 1)$ in the light of the non-robustness of the OLS and the robustness of the TLS. Redefining the dependent variable may improve the robustness of the two-stage estimator through the reduction of the influence of outliers for the errors v_t, even when the first-stage estimator is non-robust. This effect, apparent in the asymptotic representation, is confirmed in the small sample simulations. Thus, specific estimators of q could be chosen to enhance robustness, although this is not the approach of this paper.

7 Conclusion

In this paper, we propose a new technique of variance reduction for two-stage estimation procedures based on the reduced-form errors in systems of linear equations. In this setting, we show that an asymptotic bias that would occur in the first-stage reduced-form estimates of the coefficients of the exogenous variables in the structural equation is integrally and exclusively transmitted to the coefficients of the same variables in the second-stage. We apply our analyses to the case of two-stage quantile regressions, for which we show that the structural approach and the fitted-value approach amount to estimating the same slope coefficients, under a natural instrumental variable assumption.

Our leading cases are the two-stage quantile regressions with random regressors, $\alpha-$mixing and non identically distributed error terms, where the first stage is implemented with least-squares or trimmed-least-squares estimators. Our reformulation of the dependent variable introduces a trade-off between an asymptotic bias on the intercept of the equation of interest on the one hand, and the variance reduction of the slope estimator on the other hand. The reformulated variable is a weighed mean of the initial dependent variable and of its fitted-value obntained from a reduced-form estimate.

We derive the asymptotic normality and the asymptotic variance-covariance matrix of the estimator. Then, we propose to perform variance reduction for the slope coefficients of interest by estimating variance-minimising reformulation weights. Our simulation results show massive efficiency gains. In particular, our new technique alleviates the sometimes poor efficiency of quantile regressions.

Let us now discuss a practical procedure for two-stage quantile regressions. There are two basic principles. First, the first-stage estimators should be carefully selected so as to preserve efficiency, robustness or other desired properties. Our simulation results suggest that OLS should perform well under normality, while trimmed least-square should be more accurate and more robust for heavy tails or asymmetric error distributions. Second, one should reformulate the dependent variable as proposed, in such a way that the selected variance criterion is minimised.

For example, the computation steps for trimmed least-squares plus quantile regressions are: (1) trimmed least-squares for the reduced-form equation and the ancillary equations, (2) calculus of the fitted-values for the endogenous regressors of the structural equation, (3) preliminary quantile regression of the structural equation where the endogenous regressors are substituted with fitted values, (4) estimation 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 of the structural equation incorporating the fitted-value of the endogenous regressors, (7) final quantile regression of the structural equation.

As a word of conclusion, let us mention that our approach could be generalised to other first-stage and second-stage estimators, and different reformulations of the dependent variable, thus suggesting diverse variance reduction techniques. However, it is not as obvious to generalise as one may think. For our method to work, some invariance properties must be satisfied that allow for the coordination of the reformulation of the dependent variable with the error restrictions. With quantile regressions convenient invariance properties to monotone transformations can be mobilised. For other methods complications may occur.

References

[1] Abadie, A., J. Angrist and G. Imbens (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica,* 70, 91-117.

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

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

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

[5] Arias, O., K.F. Hallock and W. Sosa-Escudero (2001), "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data," *Empirical Economics*, 26, 7-40.

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

[7] Chen, L.-A. and S. Portnoy (1996), "Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Structural Equation Models," *Commun. Statist.-Theory Meth.*, 25, 1005-1032.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[25] Kim, T. and C. Muller (2006), "Bias Transmission and Variance Reduction in Two-Stage Estimation," *Working Paper IVIE.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Proofs

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

$$\sup_{t\leq T, T>1}\left[E\left(\sup_{\zeta_1:\|\zeta_1-\zeta\|<L_1}|m^*_i(w_t,\zeta_1) - m^*_i(w_t,\zeta)|^p\right)\right]^{1/p} \leq CL_1^{\psi} \tag{12}$$

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

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

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

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

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

which implies

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

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

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

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

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

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

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

$$
\sup_{||\zeta_1-\zeta_2||\leq L_T^*}||V_T^*(\zeta_1)-V_T^*(\zeta_2)||=o_p(1). \tag{13}
$$

We now introduce a factor $T^{-1/2}$ that weighs the contribution of the first stage estimator in the kernel of the empirical process. For this, we choose $L_T^* = T^{1/2}L$

for a fixed positive number L. Let $V_T(\Delta) = M_T(\Delta) - E(M_T(\Delta))$, where $M_T(\Delta) = T^{-1/2}\sum_{t=1}^{T} m(w_t,\Delta)$, $m(w_t,\Delta) = x_t\psi_\theta(qv_t - T^{-1/2}x_t'\Delta)$, and Δ is a $K\times 1$ vector. Since $V_T^*(\zeta) = V_T(T^{1/2}\zeta)$, by defining $\Delta_1 = T^{1/2}\zeta_1$ and $\Delta_2 = T^{1/2}\zeta_2$, the result in (13) becomes

$$\sup_{||\Delta_1-\Delta_2||\le L} ||V_T(\Delta_1) - V_T(\Delta_2)|| = o_p(1). \tag{14}$$

Setting $\Delta_1 = \Delta$ and $\Delta_2 = 0$ in (14), yields

$$\sup_{\|\Delta\|<L} \|M_T(\Delta) - M_T(0) - \{EM_T(\Delta) - EM_T(0)\}\| = o_p(1). \tag{15}$$

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

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

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

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

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

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

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

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 $E(M_T(\Delta)) - E(M_T(0)) \to -q^{-1}Q_0\Delta$. QED.

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

$$M_T(\hat{\Delta}_0) = M_T(0) - q^{-1}Q_0\hat{\Delta}_0 + o_p(1) \tag{16}$$

where M_T is defined in the proof of 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 $q > 0$. Therefore, under Assumptions 1, 3(iv)-(v) and 4(i), $T^{-1/2}\sum_{t=1}^{T} x_t\psi_\theta(qv_t)$ converges in distribution to a normal random variable by the CLT in Theorem 5.20 of White (2001).

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[12]. In these specific cases, the reminder of the proof is as in the leading case we study, which justifies the (b) part of the proposition.

[12] This might be one reason why authors imposed symmetry of error terms, as in Chen (1988) and Chen and Portnoy (1996).

Therefore, we have

$$M_T(\hat{\Delta}_0) = O_p(1). \tag{17}$$

These preliminary results will be used to substitute $M_T(\hat{\Delta}_0)$ in (19) below with standardised first-stage estimators and yield simplications. We now deal with the second stage empirical process.

(b) Linearization of the emprical process of the second stage: For this, we define $\hat{\Delta}_1(\delta) = H(\hat{\Pi})\delta + \hat{\Delta}_0 = H(\hat{\Pi})\delta - (1-q)T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) + T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0$ for $||\delta|| \leq L$, where $\delta \in R^{G+K_1}$ for some $L > 0$. Later, δ will be replace by the second-stage standardised estimator, $T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$. Using Assumption 2 and Lemma 1, it is straightforward to show that

$$\sup_{||\delta|| \leq L} ||M_T(\hat{\Delta}_1(\delta)) - M_T(0) + q^{-1}Q_0\hat{\Delta}_1(\delta)|| = o_p(1) \tag{18}$$

In order to regroup the first-stage and second-stage estimators, we need one more result of stochastic equicontinuity. For this, we define $\tilde{M}_T(\delta) = H(\hat{\Pi})' M_T(\hat{\Delta}_1(\delta))$ and $||H(\hat{\Pi})||^2 = tr(H(\hat{\Pi})H(\hat{\Pi})')$, which is $O_p(1)$ since $\hat{\Pi}$ converges to $\Pi_0 + B_\Pi$ that is finite.

We now use the argument between (A.7) and (A.8) in Powell (1983) to show that (17) and (18) imply that for some finite $L_2 > 0$:

$$\sup_{||\delta|| \leq L_2} ||\tilde{M}_T(\delta) - H(\Pi_0^*)' M_T(\hat{\Delta}_0) + q^{-1}Q_{zz}^*\delta|| = o_p(1) \tag{19}$$

where $Q_{zz}^* = H(\Pi_0^*)' Q_0 H(\Pi_0^*)$. Powell's argument is the following. Since $||H(\hat{\Pi})||^2 = O_p(1)$ and $||H(\hat{\Pi}) - H(\Pi_0^*)|| = o_p(1)$, we have

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

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

(c) Characterization of B_α: Next, we define $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$ in which B_α is a constant vector *defined* such as to achieve the correspondance of the FOCs of the second stage $T^{1/2}\tilde{M}_T(\hat{\Delta})$ in order to obtain

$$\tilde{M}_T(\hat{\Delta}) = o_p(1). \tag{20}$$

Moreover, if we can show $\hat{\Delta} = O_p(1)$, we shall be able to plug $\hat{\Delta}$ into (19) in place of δ and, by using (20), and cancelling the term $\tilde{M}_T(0)$ by using the second-stage FOCs, we would obtain $-H(\Pi_0^*)'M_T(\hat{\Delta}_0) + q^{-1}Q_{zz}^*\hat{\Delta} = o_p(1)$. This will deliver the asymptotic representation of $T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha)$ in the proposition.

First we search for the formula of B_α that would identify $T^{1/2}\tilde{M}_T(\hat{\Delta})$ and the second-stage FOCs. Note that

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

where

$$\begin{aligned}
\hat{A}_t &= -qx_t'\pi_0 + x_t'H(\hat{\Pi})\alpha_0 - (1-q)x_t'\pi_0 - x_t'\hat{\Pi}\gamma_0 + x_t'\Pi_0\gamma_0 \\
&= x_t'H(\hat{\Pi})\alpha_0 - x_t'H(\Pi_0)\alpha_0 + x_t'\Pi_0\gamma_0 - x_t'\hat{\Pi}\gamma_0,
\end{aligned}$$

because $\pi_0 = H(\Pi_0)\alpha_0$. Regrouping the terms involving the difference between Π_0 and $\hat{\Pi}$ and

$$\hat{B}_t = x_t'[H(\hat{\Pi})B_\alpha - (1-q)B_\pi + B_\Pi\gamma_0]$$

isolates the bias terms. $\hat{A}_t = 0$ because $x_t'H(\hat{\Pi})\alpha_0 = x_{1t}'\beta_0 + x_t'\hat{\Pi}\gamma_0$ and $x_t'H(\Pi_0)\alpha_0 = x_{1t}'\beta_0 + x_t'\Pi_0\gamma_0$. Furthermore, if we define B_α such as to impose $\hat{B}_t = 0$, then one can show that $T^{1/2}\tilde{M}_T(\hat{\Delta}) = \left[\frac{\partial S_T}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}}\right]_-$, which is the vector of left-hand-side partial derivatives of the objective function in (5) evaluated at the solution $\hat{\alpha}$. Since the second-order FOC term $\left[\frac{\partial S_T}{\partial \alpha}\Big|_{\alpha=\hat{\alpha}}\right]_-$ is at most $o_p(1)$, we obtain (20) if $\hat{B}_t = 0$. Meanwhile, $\hat{B}_t = 0$ implies that B_α is the solution to the following equation system.

$$H(\hat{\Pi})B_\alpha - (1-q)B_\pi + B_\Pi\gamma_0 = 0,$$

which can be rewritten as

$$H(\hat{\Pi})B_\alpha = b, \tag{21}$$

where $b = (1-q)B_\pi - B_\Pi\gamma_0$. Moreover, due to Assumptions 2 and 3(i), the rank of $H(\hat{\Pi})$ for a sufficiently large T is $K_1 + G$, which is equal to the dimension of B_α. This implies that there is no other solution asymptotically. The unique solution can be obtained by multiplying both sides of (21) by $H(\hat{\Pi})'Q_0$ and inverting the matrix $\hat{Q}_{zz} = H(\hat{\Pi})'Q_0H(\hat{\Pi})$:

$$B_\alpha = \hat{Q}_{zz}^{-1}H(\hat{\Pi})'Q_0\{(1-q)B_\pi - B_\Pi\gamma_0\}.$$

$\hat{Q}_{zz}$ is invertible for T large enough because of Assumptions 3(i) and 3(iii).

Let us show that $\hat{\Delta} = T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) = O_p(1)$. This will prove that B_α is the asymptotic bias of $\hat{\alpha}$. We can obtain $\hat{\Delta} = O_p(1)$ by using the argument in Lemma A.4 in Koenker and Zhao (1996). Similar arguments are in Jureckova (1977) and Hjort and Pollard (1999). To use Lemma A.4 in Koenker and Zhao (1996) and obtain $\hat{\Delta} = O_p(1)$ we need to check the following conditions:

(i) $-\delta'\tilde{M}_T(\lambda\delta) \geq -\delta'\tilde{M}_T(\delta)$ for $\lambda \geq 1$ and $||\delta|| \leq L_3$ for some $L_3 > 0$.

(ii) $||H(\Pi_0^*)'M_T(\hat{\Delta}_0)|| = O_p(1)$,

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

(iv) Q^*_{zz} is positive definite.

Condition (i) is obtained by noticing that function $h(\lambda) = \sum_{t=1}^T \rho_\theta(qv_t - T^{-1/2}x_t' H(\hat{\Pi})\delta\lambda - T^{-1/2}x_t'\hat{\Delta}_0)$ is convex in λ, and therefore that its gradient, $-\delta'\tilde{M}_T(\lambda\delta)$ is non-decreasing in λ. Condition (ii) stems from (17). Condition (iii) results from the first-order conditions of the second stage, as we discussed above. Finally, condition (iv) is ensured by Assumptions 3(i) and 3(iii). Hence, by Lemma A.4 in Koenker and Zhao (1996), we have

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

(d) Inversion: Therefore, we can plug $\hat{\Delta}$ into (19) in place of δ, and using the result in (20) gives

$$\begin{aligned} q^{-1}Q^*_{zz}\hat{\Delta} &= H(\Pi^*_0)' M_T(\hat{\Delta}_0) + o_p(1), &\quad (23)\\ &= H(\Pi^*_0)' M_T(0) - H(\Pi^*_0)' q^{-1} Q_0 \hat{\Delta}_0 + o_p(1) &\quad (24)\end{aligned}$$

where the second equality comes from (16). By plugging the definition of $\hat{\Delta}_0$ and inverting $q^{-1}Q^*_{zz}$, we obtain

$$\begin{aligned} T^{1/2}(\hat{\alpha} - \alpha_0 - B_\alpha) &= Q^{*-1}_{zz} H(\Pi^*_0)' \{T^{-1/2}\sum_{t=1}^T x_t q \psi_\theta(v_t) \\ &\quad + (1-q)Q_0 T^{1/2}(\hat{\pi} - \pi_0 - B_\pi) \\ &\quad - Q_0 T^{1/2}(\hat{\Pi} - \Pi_0 - B_\Pi)\gamma_0\} + o_p(1). \end{aligned}$$

QED.

Proof of Proposition 2:

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 $AH(\Pi^*_0) = RQ_0H(\Pi^*_0) = I_{(G+K1)}$ by definition of R. It follows that the first column of $AH(\Pi^*_0)$ is a', owing to the arrangement of elements in $H(\Pi_0^*)$. Indeed, the first column of $AH(\Pi_0^*)$ can be calculated by matrix multiplying successively the lines of A by the first column of $H(\Pi_0^*)$, which is $(1, 0'_{K-1})'$. So, only the first element of each line of A is kept and stored in the column. Consequently, since the first column of $I_{(G+K_1)}$ is $(1, 0, \ldots, 0)'$, we have $RQ_1 = \begin{bmatrix} 1 \\ 0_{K_1+G-1} \end{bmatrix}$. The proof can be similarly extended to other columns with $Q_0 = [Q_j, Q_{-j}]$, where Q_j corresponds to the first j columns f Q_0 and Q_{-j} to the other columns of $Q_0, 1 \leq j \leq K_1$. In that case, we obtain $RQ_j = \begin{bmatrix} I_j \\ 0_{K_1+G-j} \end{bmatrix}$. Therefore, in the situations where the bias term $(1-q)B_\pi - B_\Pi\gamma_0$ is restricted to the first K_1 components, we obtain the result of the Proposition. QED.

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

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

Since x'_t, u_t, v_t are α-mixing by assumption, and S_t is a measurable function of x'_t, u_t, v_t, it follows that S_t is also α-mixing. Next, $E(S_t) = 0$ by Assumptions 3(iv) and 4(ii). Finally, Assumption 4(i) provides all the moment conditions necessary to invoke Theorem 5.20 of White (2001). Hence, we have:

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

which implies the result. QED.

Calculus of q*:

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

$q^{*} = -\frac{b}{a} = \frac{E(v_t^{*}u_t^{*}) - f(0)^{-1}E(\psi_\theta(v_t)u_t^{*})}{f(0)^{-2}\theta(1-\theta) + E(v_t^{*2}) - 2f(0)^{-1}E(\psi_\theta(v_t)v_t^{*})}$. The calculus for the TLS case is similar.

Appendix B: Simulation Design

We base our simulations on the simplest possible model: a simultaneous equation system with two simple equations. The first equation, which is the equation of interest, contains two endogenous variables and two exogenous variables including a constant. Four exogenous variables are present in the whole system. The structural simultaneous equation system can be written $B \begin{bmatrix} y_t \\ Y_t \end{bmatrix} + \Gamma x_t = U_t$, where $\begin{bmatrix} y_t \\ Y_t \end{bmatrix}$ is a 2×1 vector of endogenous variables, x_t is a 4×1 vector of exogenous variables with the first element equal to one. The error term U_t is a 2×1 vector of error terms. We specify the structural parameters as follows: $B = \begin{bmatrix} 1 & -0.5 \\ -0.7 & 1 \end{bmatrix}$ and $\Gamma = \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & 0.2 \end{bmatrix}$. The system is over-identified by the exclusion restrictions $\Gamma_{13} = \Gamma_{14} = \Gamma_{22} = 0$. Moreover, $\begin{bmatrix} v & V \end{bmatrix} = U(B')^{-1}$. 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 attenuated effects are chosen for the cross effects of the two endogenous variables so that the endogeneity 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 correlation coefficient -0.1 and x_{5t} is generated from a distribution N$(0,1)$ independently of other random variables and errors. Because we assume that x_{5t} is independent of w_t and w_t is iid, $F_{v^e}^{-1}(\theta) = \sigma(x_{5t})F_w^{-1}(\theta)$, where $F_w^{-1}(\theta)$ is the inverse cumulative function of w_t evaluated at θ. The scale factor is $\sigma(x_{5t}) = 1 + \delta x_{5t}$. We choose $\delta = 0.05$ under heteroscedasticity and $\delta = 0$ under iid. The errors V_j are generated in the same way, albeit without heteroscedasticity. Then, we draw the second to fourth columns in X from the normal distribution with mean $(0.5, 1, -0.1)'$, variances normalised 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 $X, [\begin{array}{cc} v & V \end{array}]$ and $[\begin{array}{cc} \pi_0 & \Pi_0 \end{array}] = -\Gamma'(B')^{-1}$, we generate the endogenous variables $[\begin{array}{cc} y & Y \end{array}]$ by using the reduced-form equation: $[\begin{array}{cc} y & Y \end{array}] = X[\begin{array}{cc} \pi_0 & \Pi_0 \end{array}] + [\begin{array}{cc} v & V \end{array}]$