

# A Robust Test of Exogeneity Based on Quantile Regressions

Tae-Hwan Kim  
Christophe Muller

# A Robust Test of Exogeneity Based on Quantile Regressions

Tae-Hwan Kim<sup>a</sup> and Christophe Muller<sup>b,\*</sup>

<sup>a</sup>School of Economics, Yonsei University, Seoul 120-749, Korea.

Tel.: +82-2-2123-5461; fax: +82-2-2123-8638.

tae-hwan.kim@yonsei.ac.kr

<sup>b</sup>Aix-Marseille Univ, CNRS, EHESS, Ecole Centrale, Greqam, Marseille, France.

14, avenue Jules Ferry, F-13621 Aix-en-Provence Cedex, France.

christophe.muller@univ-amu.fr

April 2017

**Abstract:** In this paper, we propose a robust test of exogeneity. The test statistics is constructed from quantile regression estimators, which are robust to heavy tails of errors. We derive the asymptotic distribution of the test statistic under the null hypothesis of exogeneity at a given quantile. Then, the finite sample properties of the test are investigated through Monte Carlo simulations that exhibit not only good size and power properties, but also good robustness to outliers.

**Key words:** regression quantile, endogeneity, two-stage estimation, Hausman test.

**JEL codes:** C21.

\*The first author acknowledges financial support from the National Research Foundation of Korea – a grant funded by the Korean Government (NRF-2009-327-B00088) and from the Aix-Marseille School of Economics. The second author acknowledges financial support from the A\*MIDEX project (n° ANR-11-IDEX-0001-02) funded by the « Investissements d’Avenir » French Government program, managed by the French National Research Agency (ANR). We are grateful for comments by participants in the 2013 French Econometrics Conference and the 2014 EEA conference in Toulouse.

# 1 Introduction

Endogeneity issues are pervasive in empirical estimation of econometric models. For example, consider a typical wage equation where the logarithm of the wage rate of a worker is linearly explained by education level and some other explanatory factors. The latter factors are often considered to be independent of the error term. In contrast, the independence of the education variable and the error is generally disputed, for example because some unobservable genetic ability may be simultaneously related to both wage and education. In that case, the model may be subject to the endogeneity problem.

The typical Hausman test of endogeneity in linear models is based on comparing OLS estimates with 2SLS estimates (Hausman, 1978). Developments in Hausman-type tests used for endogeneity analysis have attracted interest in the recent literature.<sup>1</sup> However, with the typical exogenous tests in the literature, small perturbations of the distribution of the error terms in the model can have arbitrary large effects on the power and on the asymptotic level of these tests. This shortcoming stems from test statistics based on estimators that have unbounded influence functions. We contribute to this literature by exploring for this kind of test the use of quantile regressions, which are robust to heavy tail errors.

Specific kinds of robustness properties have already been the object of investigation in the literature. On the one hand, exogeneity tests that are robust to heteroskedasticity have been proposed (Wooldridge, 1995), and the robustness of exogeneity tests to the weakness of some instruments and to the degree of endogeneity of instruments have been studied (Tchatoka and Dufour, 2008). However, to our knowledge, there are no precise results for robustness of exogeneity tests to outliers.

On the other hand, another literature examines the general ‘robustification’ of tests, by basing them on statistics that are robust to heavy tails, notably for M-estimators (Peracchi, 1991). We follow this suggestion by using quantile regression, which is an estimation method robust to outliers, and applying them to exogeneity tests. To the exception of Chernozhukov and Hansen (2006)’s test, which we discuss below, no exogeneity test based on bounded-influence statistics seems to be available. We close this gap.

The issue of endogeneity in the context of quantile regression has long been recognized, and many techniques to deal with this issue have been proposed.<sup>2</sup> However, not much attention has been paid to the issue of testing for the presence of endogeneity in conditional regression models by using quantile regression. In this paper, the use of quantile regression for this is motivated by the robustness of this method when the error terms exhibit a heavy tail. Chernozhukov and Hansen (2006) propose an exogeneity test for instrumental quantile regression based on approximating the orthogonality of transformed structural errors and regressors (IVQR estimator). **to be more precise about quant C&H do** A drawback of their approach is that the computation times of the test statistic and of its critical value may be huge for more than a few endogenous regressors. On the other hand, our proposed method can accommodate as many endogenous regressors as possible, and is free from such a computational problem.

Robustness estimators are sometimes justified by non-normality since in the Gaussian case the

---

<sup>1</sup>Hahn and Hausman (2002), Butler (2000), Chmelarova and Hill (2010), Lavergne and Ngimkeu (2011), Lee and Okui (2012).

<sup>2</sup>To name just a few, see Amemiya (1982), Powell (1983), Chen and Portnoy (1996), Kemp (1999), Sakata (2007), Arias et. al. (2001), Garcia et. al. (2001), Chen, Linton, and van Keilegem (2003), Hong and Tamer (2003), Kim and Muller (2004, 2012), Chernozhukov and Hansen (2005, 2006, 2008), Ma and Koenker (2006), Horowitz and Lee (2007), and Lee (2007).

non-robust OLS estimates in the linear model is the maximum likelihood estimator and outliers are rare. The issue of weak endogeneity in linear simultaneous systems is also sometimes seen as originated from non-normality. As it happens, the use of quantile regressions can also be motivated by non-Gaussian errors since they may be preferable in that case to least-square estimators that are efficient exclusively under normality. As a result, in non-Gaussian settings with endogeneity issues, quantile regression emerges as a robust alternative to LS-based estimation methods, even when the main interest is in the central tendency of the response variable.

We use the fitted-value approach of quantile regression under endogeneity, which is inexpensive to compute, even with several endogenous variables. However, in this approach, which corresponds to the independence of the reduced-form errors from regressors, the slope coefficients of the structural model must be the same for different quantiles<sup>3</sup>. Nevertheless, using quantile regression is still useful even when the slope coefficients do not vary across quantiles, if the errors exhibit heavy tails. Indeed, quantile regression can provide a protection against outliers in the errors. As a matter of fact, only one quantile can be selected to construct the test, which makes the variation of coefficients across quantiles irrelevant.

In this paper, we propose a feasible and robust exogeneity test based on quantile regression, easy to compute with many endogenous regressors. Our test statistic is based on a quadratic distance between two estimators, which are both robust to outliers. The first estimator is the standard quantile regression estimator, which is consistent only under the null hypothesis of exogeneity at the considered quantile. The second estimator is the double-stage quantile estimator (DSQR) developed in Kim and Muller (2004), which remains consistent regardless of the presence of endogeneity. Note, as shown in Kim and Muller (2004), that the two estimators converge in probability to the same limit, despite their semi-parametric restrictions being respectively imposed on the structural and reduced-form errors. This feature stems from applying the typical results of the fitted-value approach, as discussed in Amemiya (1982) and Powell (1983), and exhibited below. Under exogeneity, quantile regression and DSQR estimate the same thing, which corresponds to a given quantile of the structural error term.

We present the model and discuss its estimation in Section 2. In Section 3, we propose and analyze the test statistic and we derive its asymptotic distribution. In Section 4, we study the finite sample properties of our test using Monte Carlo simulations. Finally, Section 5 concludes.

## 2 The Model and the Estimation Method

We are interested in the parameter ( $\alpha_0$ ) in the following structural equation for  $T$  observations:

$$\begin{aligned} y_t &= x'_{1t}\beta_0 + Y'_t\gamma_0 + u_t \\ &= Z'_t\alpha_0 + u_t, \end{aligned} \tag{1}$$

where  $[y_t, Y'_t]$  is a  $(G + 1)$ -rows vector of possible endogenous variables,  $x'_{1t}$  is a  $K_1$ -rows vector of exogenous variables,  $Z_t = [x'_{1t}, Y'_t]'$ ,  $\alpha_0 = [\beta'_0, \gamma'_0]'$  and  $u_t$  is an error term. We denote by  $x'_{2t}$  the row vector of  $K_2 (= K - K_1)$  exogenous variables absent from (1). The first element of  $x_{1t}$  is set to be one to allow for an intercept in the model.

Estimating  $\alpha_0$  for the  $\theta^{th}$ -conditional quantile of  $y_t$ , for a given  $\theta \in (0, 1)$ , can be achieved through the following minimization program:

$$\min_{\alpha} \sum_{t=1}^T \rho_{\theta}(y_t - Z'_t\alpha) \tag{2}$$

---

<sup>3</sup>Except if one allows for inconsistencies for some coefficients (Kim and Muller, 2016).

and  $\rho_\theta(z) = z\psi_\theta(z)$  where  $\psi_\theta(z) = \theta - 1_{[z \leq 0]}$  and  $1_{[\cdot]}$  is the Kronecker index. We denote the solution of (2),  $\tilde{\alpha}$ , the ‘one-stage quantile regression estimator’ for  $\alpha_0$ . The one-stage estimator  $\tilde{\alpha}$  is consistent if the following zero conditional expectation condition holds:

$$E(\psi_\theta(u_t)|Z_t) = 0. \quad (3)$$

This condition is the assumption that zero is the given  $\theta^{th}$ -quantile of the conditional distribution of  $u_t$ . It identifies the coefficients of the model once a given quantile index  $\theta \in (0, 1)$  has been chosen. Even though the slope coefficients in (1) may not vary across different quantiles, the intercept coefficient, which is the first element of  $\beta_0$ , should vary with  $\theta$  to satisfy (3) in the absence of endogeneity. However, the condition (3) is violated if there is endogeneity in  $Y_t$ , which can actually be defined as corresponding to  $E(\psi_\theta(u_t)|Z_t) \neq 0$ . In this case,  $\tilde{\alpha}$  is inconsistent. As it happens, function  $\psi_\theta$ , once normalised and removing the factor  $x_t$  from the FOCs, can also be seen as characterizing the influence function of the quantile regression estimator.<sup>4</sup> As a matter of fact, its normalized gross error sensitivity is equal to  $\gamma^* = \max\{\theta, 1 - \theta\}$ . Since it is finite, the quantile estimator is B-robust, for fixed values of the regressor  $x_t$ . The fact that function  $\psi$  is bounded between  $\theta$  and  $1 - \theta$  ensures that no outlier in the error terms can have an exaggerated impact on estimation. Exploiting this property, we develop a procedure to test robustly for exogeneity in  $Y_t$  at a given quantile  $\theta$ . The notion of exogeneity of interest is the independence of  $u_t$  and  $Z_t$ . As usual for exogeneity tests, we only formally test for a consequence of exogeneity, which is here the orthogonality of  $\psi_\theta(u_t)$  and  $Y_t$ , for any given  $\theta$ .

We assume that  $Y_t$  can be linearly predicted from the exogenous variables:

$$Y_t' = x_t' \Pi_0 + V_t', \quad (4)$$

where  $x_t' = [x_{1t}', x_{2t}']$  is a  $K$ -rows vector,  $\Pi_0$  is a  $K \times G$  matrix of unknown parameters, and  $V_t'$  is a  $G$ -rows vector of unknown error terms. By assumption, the first element of  $x_{1t}$  is 1 to allow for an intercept in the model. Using (1) and (4),  $y_t$  can also be expressed as:

$$y_t = x_t' \pi_0 + v_t, \quad (5)$$

where

$$\begin{aligned} \pi_0 &= H(\Pi_0)\alpha_0 \text{ with } H(\Pi_0) = \left[ \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}, \Pi_0 \right] \\ \text{and } v_t &= u_t + V_t' \gamma_0. \end{aligned} \quad (6)$$

As mentioned before, our test statistic is based on the double-stage quantile regression in Kim and Muller (2004), which is described below and is another robust estimation method with bounded influence function. This estimator has here several advantages over other approaches. First, the calculus involved in simultaneously comparing the asymptotic representations of the two considered estimators is tractable in that case, in a setting familiar to most applied researchers. Second, both estimators can avoid the need for grid search and is free from the curse of dimensionality, which would otherwise restrict the empirical analyses to models including only a few endogenous variables. Third, and most importantly, both estimators are robust to heavy tails of errors.

---

<sup>4</sup>Let be an estimator  $\nu_T$  of a parameter  $\nu$ , which is defined by the moment conditions  $\sum_{t=1}^T x_t \psi(y_t - x_t' \nu) = 0$ , where  $\psi$  is a non-constant differentiable real function. Then, its influence function corresponds to  $IC = \frac{x_t \psi}{-\int \frac{\partial \psi}{\partial \nu} F(dy)}$ , where  $F$  is the cdf of  $y$ . See for example Huber (1981).

Equations (4) and (5) are the basis of the first-stage estimation that yields the consistent estimators  $\hat{\pi}$ ,  $\hat{\Pi}$ , respectively, of  $\pi_0$ , and  $\Pi_0$ . Specifically,  $\hat{\pi}$  and  $\hat{\Pi}_j$  (the  $j^{\text{th}}$  column of  $\hat{\Pi}$ ;  $j = 1, \dots, G$ ) are the first-stage estimators obtained by:

$$\min_{\pi} \sum_{t=1}^T \rho_{\theta}(y_t - x_t' \pi) \quad (7)$$

$$\text{and} \quad \min_{\Pi_j} \sum_{t=1}^T \rho_{\theta}(Y_{jt} - x_t' \Pi_j), \quad (8)$$

where  $\pi$  and  $\Pi_j$  are  $K \times 1$  vectors and  $Y_{jt}$  is the  $(j, t)^{\text{th}}$  element of  $Y$ . Estimating  $\pi$  will be useful later for calculating an estimate of the residual  $\hat{v}_t$ , which is a component of the estimator of the variance-covariance matrix that occurs in the formula of the test statistics. Based on these first-stage estimators, the second-stage estimator  $\hat{\alpha}$  for  $\alpha_0$  is denoted as the double-stage quantile estimator (DSQR) and is obtained by:

$$\min_{\alpha} \sum_{t=1}^T \rho_{\theta}(y_t - x_t' H(\hat{\Pi}) \alpha).$$

We need the two following regularity assumptions in order to derive the asymptotic distributions of  $\tilde{\alpha}$  and  $\hat{\alpha}$ . Let  $h(\cdot|x)$ ,  $f(\cdot|x)$ , and  $g_j(\cdot|x)$  be the conditional densities, respectively, for  $u_t$ ,  $v_t$ , and  $V_{jt}$ .

**Assumption 1** *The sequence  $\{(x_t', u_t, V_t')\}$  is independent and identically distributed.*

Assumption 1 facilitates the presentation of our results. It arises, for example, when the sources of uncertainty in the data come from randomly sampling the observations. Assumption 1 could be relaxed to allow for serial correlation and heteroskedasticity.

**Assumption 2** (i)  $E(\|x_t\|^3) < \infty$  and  $E(\|Y_t\|^3) < \infty$ , where  $\|a\| = (a'a)^{1/2}$ .

(ii)  $H(\Pi_0)$  is of full column rank.

(iii) *There is no hetero-altitudinality:  $h(\cdot|x) = h(\cdot)$ ,  $f(\cdot|x) = f(\cdot)$  and  $g_j(\cdot|x) = g_j(\cdot)$ , where  $h(\cdot)$ ,  $f(\cdot)$  and  $g_j(\cdot)$  are assumed to be continuous. Moreover, all densities are positive when evaluated at zero:  $h(0) > 0$ ,  $f(0) > 0$ , and  $g_j(0) > 0$ .*

(iv) *There exist constants  $\lambda_h$ ,  $\lambda_f$ , and  $\lambda_j$  such that  $h(\cdot) < \lambda_h$ ,  $f(\cdot) < \lambda_f$ , and  $g_j(\cdot) < \lambda_j$ .*

(v) *The matrices  $Q_x = E(x_t x_t')$  and  $Q_z = E(Z_t Z_t')$  are finite and positive definite.*

(vi)  $E\{\psi_{\theta}(v_t) | x_t\} = 0$  and  $E\{\psi_{\theta}(V_{jt}) | x_t\} = 0$  ( $j = 1, \dots, G$ ), for a given quantile  $\theta$ .

Assumption 2(i) is necessary for obtaining the stochastic equicontinuity of the relevant empirical process in the dependent case which is used for deriving the asymptotic representation of our estimators. We also use it to limit the asymptotic variance-covariance matrix of the estimators. Assumption 2(ii) is the usual identification condition for simultaneous equations models. Assumption 2(iii) allows us to simplify the asymptotic variance-covariance matrix of the double-stage quantile regression estimator. However, combined with imposing that the slope coefficients do not vary across quantiles, Assumption 2(iii) becomes close to the independence between the error terms and the exogenous regressors. It could be relaxed at the cost of using more complicated formulae for the asymptotic variance-covariance matrix of the DSQR estimator, and, as a consequence, for the test

statistics and its distribution. Assumption 2(iv) limiting the densities simplifies the demonstration of convergence to zero for the remainder terms in the calculus of the asymptotic representation. Assumption 2(v) is the counterpart of the usual condition for OLS under which the sample second moment matrix of the regressor vectors converges towards a finite positive definite matrix. It ensures that  $E(x_t Y_t) \neq 0$  and  $E(Z_t Y_t) \neq 0$ . Finally, Assumption 2(vi) imposes that zero is the  $\theta^{th}$ -quantile of the conditional distributions of  $v_t$  and of each  $V_{jt}$ .<sup>5</sup> It identifies the coefficients of the model. For example  $\theta = 0.5$  would correspond to LAD estimator and a central tendency. When error terms and regressors are independent, Assumption 2(vi) would be satisfied for any  $\theta$ . In that case, the corresponding intercept coefficient should vary with  $\theta$  to satisfy the conditions, given that the slope coefficients are fixed across quantiles. Note that the restrictions in Assumption 2(vi) are applied to the reduced-form errors ( $v_t$  and  $V_{jt}$ ) instead of the structural error ( $u_t$ ). This is typical of the fitted-value approach, originally proposed by Amemiya (1982) for LAD estimators, and still allows for consistent estimation of the structural model. It is also possible to stick to one quantile only, which is the approach of this paper whose emphasis is on robustness.

The asymptotic representation of the quantile regression estimator  $\tilde{\alpha}$  is well known:

$$T^{1/2}(\tilde{\alpha} - \alpha_0) = Q_z^{-1} T^{-1/2} \sum_{t=1}^T Z_t \epsilon_{1t} + o_p(1), \quad (9)$$

where  $\epsilon_{1t} = h(0)^{-1} \psi_\theta(u_t)$ . From (9), it is clear that  $\tilde{\alpha}$  is consistent if  $T^{-1} \sum_{t=1}^T Z_t \epsilon_{1t}$  vanishes in probability. Given that the probability limit  $E(Z_t \epsilon_{1t})$  is zero in the absence of endogeneity, we have in that case

$$T^{1/2}(\tilde{\alpha} - \alpha_0) \xrightarrow{d} N(0, \sigma_{11} Q_z^{-1}),$$

where  $\sigma_{11} = E(\epsilon_{1t}^2) = h(0)^{-2} \theta(1 - \theta)$  and  $Q_z = E(Z_t Z_t')$ . The covariance term  $\sigma_{11} Q_z^{-1}$  can be consistently estimated by  $\hat{\sigma}_{11} \hat{Q}_z^{-1}$ , where  $\hat{Q}_z = T^{-1} \sum_{t=1}^T Z_t Z_t'$  and  $\hat{\sigma}_{11} = T^{-1/2} \sum_{t=1}^T \hat{\epsilon}_{1t}^2 = \hat{h}(0)^{-2} \theta(1 - \theta)$  with  $\hat{\epsilon}_{1t} = \hat{h}(0)^{-1} \psi_\theta(\hat{u}_t)$ ,  $\hat{u}_t = y_t - Z_t \hat{\alpha}$ . Here,  $\hat{h}(0)$  can be any consistent kernel-type non-parametric estimator of density  $h$ , calculated at zero.

A corresponding result can be obtained for the double-stage estimator  $\hat{\alpha}$  (see Kim and Muller, 2004):

$$T^{1/2}(\hat{\alpha} - \alpha_0) = Q_{zz}^{-1} H(\Pi_0)' T^{-1/2} \sum_{t=1}^T x_t \epsilon_{2t} + o_p(1), \quad (10)$$

where  $Q_{zz} = H(\Pi_0)' Q_x H(\Pi_0)$ ,  $Q_x = E(x_t x_t')$ , and  $\epsilon_{2t} = f(0)^{-1} \psi_\theta(v_t) - \sum_{i=1}^G \gamma_{0i} g_i(0)^{-1} \psi_\theta(V_{it})$ . Note that the error  $\epsilon_{2t}$  is bounded for any arbitrarily large errors  $v_t$  since  $\psi_\theta$  is bounded, which ensures robustness to outliers in errors. By ‘inverting’ the expansion (10) and applying a CLT, we have:

$$T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \sigma_{22} Q_{zz}^{-1}),$$

where  $\sigma_{22} = E(\epsilon_{2t}^2)$ . As before,  $\sigma_{22}$  and  $Q_{zz}$  can be consistently estimated as follows:  $\hat{Q}_{zz} = H(\hat{\Pi})' \hat{Q}_x H(\hat{\Pi})$  with  $\hat{Q}_x = T^{-1} \sum_{t=1}^T x_t x_t'$  and  $\hat{\sigma}_{22} = T^{-1/2} \sum_{t=1}^T \hat{\epsilon}_{2t}^2$  with  $\hat{\epsilon}_{2t} = \hat{f}(0)^{-1} \psi_\theta(\hat{v}_t) - \sum_{i=1}^G \hat{\gamma}_{0i} \hat{g}_i(0)^{-1} \psi_\theta(\hat{V}_{it})$ , where  $\hat{f}(0)$  and  $\hat{g}_i(0)$  are kernel-type estimators of  $f(0)$  and  $g_i(0)$ , respectively, and  $\hat{v}_t$  and  $\hat{V}_{it}$  are the residuals from the first-stage regressions in (7) and (8).

<sup>5</sup>Note that in the iid case, the term  $f(F^{-1}(\theta))^{-1}$  typically appears in the variance formula of a quantile estimator (Koenker and Bassett, 1978). However, due to Assumption 3(iv),  $F^{-1}(\theta)$  is now zero so that in this case, we instead have  $f(0)^{-1}$ .

### 3 The Exogeneity Test

#### 3.1 The test

The null hypothesis we consider is:

$$H_0 : \text{There is no endogeneity in the given } \theta^{th} \text{ quantile,}$$

which is equivalent to:

$$H_0 : E(\psi_\theta(u_t)|Z_t) = 0. \quad (11)$$

Given that slope quantile estimators are consistent regardless of the value of  $\theta$ , we use the slope estimators only to construct a test statistic (denoted by  $KM$ ). Moreover, the default in Hausman test procedures implemented by popular software (e.g., Stata) is to exclude the intercept in the comparison. This is seen as appropriate for models in which the constant does not have common interpretation across the models. Specifically, let  $\alpha_{0(1)}$  and  $\alpha_{0(2)}$  be the intercept and slope coefficients respectively, and let us decompose the quantile estimators  $\tilde{\alpha}$  and  $\hat{\alpha}$  accordingly; that is,  $\tilde{\alpha}' = (\tilde{\alpha}_{(1)}, \tilde{\alpha}'_{(2)})$  and  $\hat{\alpha}' = (\hat{\alpha}_{(1)}, \hat{\alpha}'_{(2)})$ . The principle driving the test is that both slope estimators  $\tilde{\alpha}_{(2)}, \hat{\alpha}_{(2)}$  for a given  $\theta$  are consistent and asymptotically normal under the null hypothesis of no endogeneity, while only the slope estimator  $\hat{\alpha}_{(2)}$  is consistent under the alternative hypothesis of endogeneity at quantile  $\theta$ . Thus, a quadratic distance between  $\tilde{\alpha}_{(2)}$  and  $\hat{\alpha}_{(2)}$  can be used to test consistently the null hypothesis of exogeneity.

If we wished to place ourselves in the original Hausman test setting (Hausman, 1978), in a strict sense,  $\tilde{\alpha}_{(2)}$  should be efficient under  $H_0$ . However, quantile regression is not generally asymptotically efficient, even under exogeneity.<sup>6</sup> As a consequence, we cannot use the difference of asymptotic variance-covariance matrices of the two estimators as equivalent to the asymptotic variance-covariance matrix of the gap  $\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}$ , as in the usual Hausman test. Indeed, we must calculate the joint covariance of the two estimators, which will be obtained from considering together their asymptotic representations. That is: we allow for inefficient estimators by dealing with the joint distribution of the estimators without invoking orthogonality conditions between the estimators. To recap, under the null hypothesis of exogeneity at the given  $\theta^{th}$  quantile, both the quantile regression and the double quantile regression converge to the same values. On the other hand, under the alternative hypothesis of endogeneity at the given  $\theta^{th}$  quantile, the gap of the two slope estimators diverges. These features ensure that our test is consistent.

We first show that the variance-covariance matrix of  $\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}$  is  $R_{(2)}C^{-1}R'_{(2)}$ , where  $R_{(2)}$  is the matrix composed of the last  $(K_1 + G - 1)$  rows in  $R = [ I_{K_1+G} : -I_{K_1+G} ]$  and  $C$  is defined in Theorem 1 below. This justifies to consider a preliminary and ancillary statistic is  $T(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)})[R_{(2)}C^{-1}R'_{(2)}]^{-1}(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)})$ , when the variance-covariance matrix is invertible. In a second stage, we deal with the estimation of the core matrix in the preliminary statistics to produce the final test statistics. Theorem 1 reports the null distribution of the preliminary statistic. The proof is provided in the Appendix.

**Theorem 1.** *Suppose that Assumptions 1 and 2 hold. Then,*

(a) *under the null hypothesis of no endogeneity at quantile  $\theta$ , we have:*

$$T(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)})[R_{(2)}C^{-1}R'_{(2)}]^{-1}(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) \xrightarrow{d} \chi^2(G),$$

---

<sup>6</sup>However, they may be efficient in some particular cases. For example, LAD regressions are efficient under errors following a Laplace distribution law.



where

$$C = \begin{bmatrix} \sigma_{11}Q_z^{-1} & \sigma_{12}Q_z^{-1}Q_{zx}H(\Pi_0)Q_{zz}^{-1} \\ \sigma_{12}Q_{zz}^{-1}H(\Pi_0)'Q'_{zx}Q_z^{-1} & \sigma_{22}Q_{zz}^{-1} \end{bmatrix}$$

and  $Q_{zx} = E(Z_t x'_t)$  and  $\sigma_{12} = E(\epsilon_{1t}\epsilon_{2t})$ .

(b) The generalized inverse can be used whenever the matrix is not full rank.

In practice,  $C$  can be replaced with a consistent estimator  $\hat{C}_T$  without affecting the limiting distribution. The following consistent estimator for  $C$  is obtained by applying the plug-in principle:

$$\hat{C}_T = \begin{bmatrix} \hat{\sigma}_{11}\hat{Q}_z^{-1} & \hat{\sigma}_{12}\hat{Q}_z^{-1}\hat{Q}_{zx}H(\hat{\Pi})\hat{Q}_{zz}^{-1} \\ \hat{\sigma}_{12}\hat{Q}_{zz}^{-1}H(\hat{\Pi})'\hat{Q}'_{zx}\hat{Q}_z^{-1} & \hat{\sigma}_{22}\hat{Q}_{zz}^{-1} \end{bmatrix},$$

where

$$\hat{Q}_{zx} = T^{-1} \sum_{t=1}^T Z_t x'_t \text{ and } \hat{\sigma}_{12} = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{1t}\hat{\epsilon}_{2t}.$$

The consistency of  $\hat{C}_T$  is stated in Lemma 1, whose proof is in the Appendix.

**Lemma 1.** *Suppose that the kernel-type density estimators  $\hat{h}(0)$ ,  $\hat{f}(0)$ , and  $\hat{g}_i(0)$  are respectively consistent for  $h(0)$ ,  $f(0)$  and  $g_i(0)$ ,  $i = 1, \dots, G$ . Then, under Assumptions 1 and 2, we have*

$$\hat{C}_T \xrightarrow{p} C.$$

**Theorem 2.** *Suppose that the kernel-type density estimators  $\hat{h}(0)$ ,  $\hat{f}(0)$ , and  $\hat{g}_i(0)$  are respectively consistent for  $h(0)$ ,  $f(0)$ , and  $g_i(0)$ , respectively. Then, under Assumptions 1 and 2, we have*

$$KM = T(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)})[R_{(2)}^{-1}\hat{C}_T^{-1}R'_{(2)}]^{-1}(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) \xrightarrow{d} \chi^2(G).$$

The result of Theorem 2 easily follows from Theorem 1 and Lemma 1. Although non-robust density estimators are still likely to yield improved results, it is advisable to choose robust density estimators. In that case, the statistic  $KM$  is robust to heavy tail errors since it is exclusively composed of robust terms. In the next section, we examine the finite-sample performance of the KM test by using Monte Carlo simulations.

## 4 Monte Carlo Simulations

The results obtained in the previous section hold in large samples. In this section, the finite sample size and power of the KM test are studied through Monte Carlo simulations. We investigate the degree of robustness of the KM test in finite samples, notably as compared to that of the standard Hausman test.

We use a simultaneous equation system composed of two equations. The first equation, which is the equation of interest, contains two endogenous variables at a given quantile  $\theta$ , and two exogenous variables, including a constant. In total, four exogenous variables are present in the whole system. The structural simultaneous equation system can be written

$$B \begin{bmatrix} y_t \\ Y_t \end{bmatrix} + \Gamma x_t = U_t, \quad (12)$$

where  $\begin{bmatrix} y_t \\ Y_t \end{bmatrix}$  is a  $2 \times 1$  vector of endogenous variables, and  $x_t$  is a  $4 \times 1$  vector of exogenous variables with the first element equal to one. The error term  $U_t = \begin{bmatrix} u_t \\ w_t \end{bmatrix}$  is a  $2 \times 1$  vector of error terms. We specify the structural parameters as follows:  $B = \begin{bmatrix} 1 & -0.3 \\ \delta & 1 \end{bmatrix}$  and  $\Gamma = \begin{bmatrix} -1 & -0.2 & 0 & 0 \\ -1 & 0 & -0.4 & -0.5 \end{bmatrix}$ . The system is overidentified by the zero restrictions  $\Gamma_{13} = \Gamma_{14} = \Gamma_{22} = 0$ .

We generate the error terms  $U_t$  using some bivariate distributions (Gaussian, Student, contaminated Gaussian). Then, we draw the second to fourth elements  $x_t$  from the normal distribution with mean  $(0.5, 1, -0.1)'$ , variances equal to 1 for normalization,  $cov(x_{2t}, x_{3t}) = 0.3$ ,  $cov(x_{2t}, x_{4t}) = 0.1$  and  $cov(x_{3t}, x_{4t}) = 0.2$ , where  $x_{2t}$ ,  $x_{3t}$  and  $x_{4t}$  are the non-constant elements of  $x_t$ . Once  $x_t$  and  $U_t$  are generated, the endogenous variables  $y_t$  and  $Y_t$  are obtained through (12). The first structural equation is

$$y_t = 0.3 Y_t + 1 + 0.2 x_{2t} + u_t, \quad (13)$$

where the presence of endogeneity depends on the  $\delta$  parameter in the second equation.

Note that if  $\delta = 0$ , there is no endogeneity at any quantile index  $\theta$  in (13). On the other hand, endogeneity at  $\theta$  occurs if  $\delta \neq 0$ . To save space, the simulations are shown only for  $\theta = 0.5$ , which corresponds to LAD estimator. Because the magnitude of  $\delta$  determines the strength of endogeneity, we can use it to analyse the empirical power of the KM test. We select a few values for  $\delta$ : 0, 0.05, 0.1, 0.15, 0.2, 0.25, and 0.3. For each of the values, we simulate the rejection probabilities by the KM test and the standard Hausman test, for the null hypothesis of exogeneity, at the 5 % significance level and based on 3,000 replications.

First, we draw the error terms  $u_t$  in (13) from the standard normal density  $N(0, 1)$ . This constitutes a benchmark for comparison since it is expected in that case that the standard Hausman test will outperform the KM test because there should be few outliers and therefore little need for robustness. The results are displayed in Tables 1(a) to 1(c) for  $T = 200, 300$  and 500 respectively. For  $\delta = 0$ , the simulated rejection probability turns out in each case to be reasonably close to the nominal 5 % level, although the test appears to be slightly undersized when the sample size is small. As expected, when it comes to power with  $\delta \neq 0$ , the Hausman test is more powerful than the KM test for any considered positive values of  $\delta$ . For example, when  $T = 300$ , the empirical power of the Hausman test is almost twice that of the KM test.

Next we turn to cases in which robustness may be needed because the data are contaminated. We use three different methods to generate the contaminated error terms  $u_t$  in (13): (i) from the student-t distribution  $t(k)$  with  $k = 1$ , which is the Cauchy distribution, (ii) from  $N(0, 1)$  as in the benchmark case, while allowing for a single outlier as in Kim and White (2004), (iii) from  $N(0, 1)$  as in the benchmark case, while allowing for a fraction of outliers. That is: a given fraction of the error terms is contaminated, as in Maronna and Yohai (1995).

The results for the Cauchy distribution case are reported in Tables 2(a)-2(c) for  $T = 200, 300$  and 500, respectively. As Table 2(a) shows, unlike the previous Gaussian case, the finite sample properties of the KM test are greatly improved, even with  $T = 200$ . The empirical power of the test rapidly converges to 100% as  $\delta$  increases. This convergence accelerates as the sample size increases, as shown in Tables 2(b) and 2(c). In contrast, still with  $T = 200$ , the Hausman test is characterized by some strange behavior in Table 2(a). Its empirical power instantly reaches its top as soon as  $\delta$  deviates from the null value of zero, and then it gradually decreases as  $\delta$  moves farther away from zero. Moreover, this peculiar phenomenon does not vanish as the sample size increases, as shown

in Tables 2(b) and 2(c).

Tables 3(a)-3(c) report the results for the single outlier case. For this, we first generate random numbers  $u_1, u_2, \dots, u_T$ . Then, we multiply the quartile of these numbers by a constant  $m = 48.62$ .<sup>7</sup> The result is then used as an outlier. The KM test turns out to be fairly robust to the outlier: Namely, the finite sample performance of the KM test is hardly changed. On the other hand, the Hausman test is severely affected with its empirical powers much lower than those of the KM test, regardless of the sample size. However, unlike the previous Cauchy distribution case, the behavior of the Hausman test remains regular with its empirical power increasing as  $\delta$  increases, although it increases at a very slow rate.

We turn to the final case with a fraction of outliers. It is natural to conjecture that the finite sample performance of each test should worsen as the contaminated fraction increases, and that the deterioration will be worse for the Hausman test than the robust KM test. To save space, we report only two cases; (i) 5% contamination in Table 4, and (ii) 20% contamination in Table 5. In the 5% contamination case, the KM test is affected, but only moderately, whereas the Hausman test is massively degraded with an empirical power not much exceeding its nominal 5% size. For example, the empirical power is only 11% for the Hausman test with  $\delta = 3$  and  $T = 300$ , while it reaches a substantial 40% for the KM test. Moreover, increasing the sample size does not seem to help much for the Hausman test, which has an empirical power with  $T = 500$  of only 16%. On the contrary, the empirical power of the KM test improves a lot, from 40% to 66%. Although, the 20% contamination case, may be deemed as less realistic, it illustrates what may happen to a non-robust test. Table 5 shows that the KM test is also somewhat affected in this extreme case, while it behaves regularly with its power increasing, as either  $\delta$  or  $T$  increases. On the other hand, the Hausman test can be hardly used as a test in this case. Its empirical power does not differ much from its 5% nominal size, and increasing the sample size does not seem to generate any improvement.

## 5 Conclusion

In this paper, we have proposed a robust test of exogeneity in linear models. The test statistics is based on a quadratic norm between a quantile regression estimator that is consistent only under exogeneity at a given conditional quantile, and another quantile regression estimator that is consistent regardless of the presence of endogeneity. The derived asymptotic null distribution of the test statistic is the usual Chi-square distribution. Monte Carlo simulations indicate that the test has excellent robustness properties, as opposed to the usual Hausman test of exogeneity.

---

<sup>7</sup>Readers are referred to Kim and White (2004) for how the value of  $m$  is determined.

## 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] 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.
- [4] Blundell, R., X. Chen and D. Kristensen (2007), “Semi-nonparametric IV estimation of shape invariant Engel curves,” *Econometrica*, 75, 1613-1669.
- [5] Butler, J.S. (2000), “Efficiency results of MLE and GMM estimation with sampling weights,” *Journal of Econometrics*, 96, 25-37.
- [6] 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, 1005-32.
- [7] 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.
- [8] Chen, X. and D. Pouzo (2009), “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals,” *Journal of Econometrics*, 152, 46-60.
- [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] Chesher, A. (2003), “Identification in nonseparable models,” *Econometrica*, 71, 1405-1441.
- [13] Chevapatrakul, T., T.-H. Kim and P. Mizen (2009), “The Taylor principle and monetary policy approaching a zero bound on nominal rates: quantile regression results for the United States and Japan,” *Journal of Money, Credit and Banking*, 41, 1705-1723.
- [14] Chmelarova, V. and R.C. Hill (2010), “The Hausman pretest estimator,” *Economics Letters*, 108, 96-99.
- [15] Garcia, J., P.J. Hernandez and A. Lopez-Nicolàs (2001), “How wide is the gap? An investigation of gender wage differences using quantile regression.” *Empirical Economics*, 26, 149-67.
- [16] Hahn, J. and J. Hausman (2002), “A New Specification Test for the Validity of Instrumental Variables,” *Econometrica*, Vol. 70, N0. 1, 163-189, January.
- [17] Hausman, J.A. (1978), “Specification tests in econometrics,” *Econometrica*, 1251-1271.

- [18] Horowitz, J.L. and S. Lee (2007), "Nonparametric instrumental variable estimation of a quantile regression model," *Econometrica*, Vol. 75, No. 4, 1191-1208.
- [19] Hong, H. and E. Tamer (2003), "Inference in censored models with endogenous regressors," *Econometrica*, 71, 905-932.
- [20] Huber, P.J. (1981), "*Robust Statistics*," John Wiley and Sons, New York.
- [21] Kemp, G. (1999), "Least absolute error difference estimation of a single equation from a simultaneous equations system." Mimeo University of Essex.
- [22] Kim, T-H. and C. Muller (2004), "Two-stage quantile regressions when the first stage is based on quantile regressions," *The Econometrics Journal*, 7, 218-231.
- [23] Kim, T-H. and C. Muller (2012), "Bias Transmission and Variance Reduction in Two-Stage Quantile Regressions," mimeo University of Aix-Marseille.
- [24] Kim, T-H. and C. Muller (2016), "A Particular Form of Non-Constant Effect in Two-Stage Quantile Regression," Working Paper University of Aix-Marseille.
- [25] Kim, T-H. and White, H. (2003), "Estimation, inference, and specification testing for possibly misspecified quantile regression," *Advances in Econometrics*, 17, 107-132.
- [26] Kim, T-H. and White, H. (2004), "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, 1, 56-73.
- [27] Lavergne, P. and P-E. Ngumkeu (2011), "A Hausman Specification Test of Conditional Moment Restrictions," January, Working Paper Simon Fraser University.
- [28] Lee, S. (2007), "Endogeneity in quantile regression models: a control function approach," *Journal of Econometrics*, 141, 1131-1158.
- [29] Lee, Y. and R. Okui (2012), "Hahn-Hausman test as a specification test," *Journal of Econometrics*, 167, 133-9.
- [30] Ma, L. and R. Koenker (2006), "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134, 471-506.
- [31] Maronna, R. and V.J. Yohai (1995), "The behavior of the Stahel-Donoho robust multivariate estimator," *Journal of the American Statistical Association*, 90, 330-341.
- [32] Peracchi, F. (1991), "Robust M-Test," *Econometric Theory*, Vol. 7(01), 69-84, March.
- [33] Powell, J. (1983), "The asymptotic normality of two-stage least absolute deviations estimators." *Econometrica*, 51, 1569-75.
- [34] Sakata, S. (2007), "Instrumental Variable Estimation Based on Conditional Median Restriction," *Journal of Econometrics*, 141, 350-382.
- [35] Tchatoka, F.S.D. and J.-M. Dufour (2008), "Instrument Endogeneity and Identification-Robust Tests: Some Analytical Results," *Journal of Statistical Inference*, 138, 2649-2661.
- [36] Wolters, M.H. (2012), "Estimating monetary policy reaction functions using quantile regressions," *Journal of Macroeconomics*, 34, 342-361.

- [37] Wooldridge, J.-M. (1995), "Score diagnostic for linear models estimated by two stage least squares," in G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan (Eds.), "Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C.R. Rao," 66-87, Vol. 7(1), pages 31-51, Jan-March, Oxford, Blackwell Ltd.

## Appendix

**Proof of Theorem 1:** Let  $\hat{\delta} = (\tilde{\alpha}', \hat{\alpha}')$  and  $\delta_0 = (\alpha'_0, \alpha'_0)'$ . Using (9) and (10), we have

$$\begin{aligned} T^{1/2}(\hat{\delta} - \delta_0) &= \begin{bmatrix} Q_z^{-1} T^{-1/2} \sum_{t=1}^T Z_t \epsilon_{1t} + o_p(1) \\ Q_{zz}^{-1} H(\Pi_0)' T^{-1/2} \sum_{t=1}^T x_t \epsilon_{2t} + o_p(1) \end{bmatrix} \\ &= DT^{-1/2} \sum_{t=1}^T S_t + o_p(1), \end{aligned} \quad (14)$$

where

$$D = \begin{bmatrix} Q_z^{-1} & 0 \\ 0 & Q_{zz}^{-1} H(\Pi_0)' \end{bmatrix} \text{ and } S_t = \begin{bmatrix} Z_t \epsilon_{1t} \\ x_t \epsilon_{2t} \end{bmatrix}. \quad (15)$$

Let us now consider (14). Vector  $S_t$  is iid by Assumption 1, and  $E(S_t) = 0$  under the null hypothesis of exogeneity at  $\theta$  and Assumption 2(vi). Hence, in order to apply the Lindeberg-Levy CLT to  $T^{-1/2} \sum_{t=1}^T S_t$ , it is sufficient to show that  $\text{var}(S_t)$  is bounded. The moment conditions on  $x_t$  and  $Y_t$  in Assumption 2(i) are sufficient for this purpose because  $\psi_\theta(\cdot)^2$  is bounded from above and all the densities evaluated at zero are positive.

Given that

$$\text{var}(S_t) = \Omega = \begin{bmatrix} \sigma_{11} Q_z & \sigma_{12} Q_{zx} \\ \sigma_{12} Q'_{zx} & \sigma_{22} Q_x \end{bmatrix}, \quad (16)$$

where  $\sigma_{ij} = \text{Cov}(\epsilon_{it}, \epsilon_{jt})$ , we have  $T^{-1/2} \sum_{t=1}^T S_t \xrightarrow{d} N(0, \Omega)$ , which implies that

$$T^{1/2}(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, C), \text{ where } C = D\Omega D'.$$

Note first that (i) the constant matrix  $D$  in (14) identical to that of the standard Hausman test statistic; and second, that (ii) the covariance matrix  $\Omega$  of  $S_t$  in (14) is also identical to the corresponding covariance matrix of the Hausman test statistic, except that the scalar terms  $\sigma_{ij}$  are different.<sup>8</sup>

Noting that  $T^{1/2}(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) = R_{(2)} T^{1/2}(\hat{\delta} - \delta_0)$ , we have

$$T^{1/2}(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) \xrightarrow{d} N(0, R_{(2)} C R'_{(2)}),$$

which, in turn, implies that

$$T(\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) [R_{(2)} C^{-1} R'_{(2)}]^{-} (\tilde{\alpha}_{(2)} - \hat{\alpha}_{(2)}) \xrightarrow{d} \chi^2(G),$$

where the generalized inverse can be used when the core matrix is not invertible.

The fact that the degrees of freedom to use in the test are equal to  $G$  can be seen from the form of the formula within brackets, above (14). This is because the (upper) rows of  $T^{1/2}(\hat{\delta} - \delta_0)$  corresponds to the components of explanatory variables in  $Z_t$  that are also in the instruments  $x_t$ . *These rows are identical, up to  $o_p(1)$ , to the (lower) rows of  $T^{1/2}(\hat{\delta} - \delta_0)$  that correspond to the same variables in  $x_t$ .* This identity occurs whether there is endogeneity at the given  $\theta$  or not.

Furthermore, the  $o_p(1)$  in these asymptotic representations are negligible as compared to a  $\chi^2(G)$  random variable, which is  $O_p(1)$  as any non-degenerate random variable.

---

<sup>8</sup>In the Hausman test specification, one would have instead:  $\sigma_{11} = \text{var}(u_t)$ ,  $\sigma_{22} = \text{var}(v_t)$ , and  $\sigma_{12} = \text{cov}(u_t, v_t)$ .

The above vector asymptotic representation can be written as

$$T^{1/2}(\hat{\delta} - \delta_0) = \left[ T^{1/2} \begin{pmatrix} Q_z^{-1} Z \epsilon_1 \\ Q_{zz}^{-1} H(\Pi_0)' x \epsilon_2 \end{pmatrix} + o_p(1) \right],$$

where  $Z \epsilon_1$  denotes the empirical mean of the  $Z_t \epsilon_{1t}$ , and  $x \epsilon_2$  denotes the empirical mean of the  $x_t \epsilon_{2t}$ .

In this expression we want to factorize a term also as to exhibit the identity (up to  $o_p(1)$ ) of the lines involving the same  $x_{1t}$ , which will cancel out when applying matrix  $R$ . Let us examine the shape of  $Q_{zz}$ .

$$\begin{aligned} Q_{zz} &= H(\Pi_0)' Q_x H(\Pi_0) = \begin{bmatrix} I_{K_1} & 0 \\ \Pi_0^{1'} & \Pi_0^{2'} \end{bmatrix} \begin{bmatrix} x_1 x_1' & x_1 x_2' \\ x_2 x_1' & x_2 x_2' \end{bmatrix} \begin{bmatrix} I_{K_1} & \Pi_0^1 \\ 0 & \Pi_0^2 \end{bmatrix} + o_p(1) \\ &= \begin{bmatrix} I_{K_1} & 0 \\ \Pi_0^{1'} & \Pi_0^{2'} \end{bmatrix} \begin{bmatrix} x_1 x_1' & x_1 x_1' \Pi_0^1 + x_1 x_2' \Pi_0^2 \\ x_2 x_1' & x_2 x_1' \Pi_0^1 + x_2 x_2' \Pi_0^2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 x_1' & x_1 x_1' \Pi_0^1 + x_1 x_2' \Pi_0^2 \\ \Pi_0^{1'} x_1 x_1' + \Pi_0^{2'} x_1 x_2' & \dots \end{bmatrix}. \end{aligned}$$

Because  $Y_t = x_{1t}' \Pi_0^1 + x_{2t}' \Pi_0^2 + V_t$  and  $E(V_t | x_t) = 0$ , we have:

$x_1 Y' = x_1 x_1' \Pi_0^1 + x_1 x_2' \Pi_0^2$ , and similar results for the other terms of matrix  $Q_{zz}$ .

We obtain

$$Q_{zz} = \begin{bmatrix} x_1 x_1' & x_1 Y' \\ Y x_1' & Y Y' \end{bmatrix} + o_p(1), \text{ which is also valid for } Q_z.$$

Therefore,  $Q_{zz}^{-1} = Q_z^{-1} + o_p(1)$ , as the inverses can be constructed for the terms of each matrix and the  $+ o_p(1)$  can be collected together.

It remains to incorporate  $H(\Pi_0)' = \begin{bmatrix} I_{K_1} & 0 \\ \Pi_0^{1'} & \Pi_0^{2'} \end{bmatrix}$  in the above asymptotic representation as a right factor of  $Q_{zz}^{-1}$ . But if we limit our attention to the  $K_1$  first lines of the empirical processes in the parenthesis, that is: to the terms in  $x_{1t}$ , we can see that these terms are not changed by applying matrix  $I_{K_1}$ . Therefore, the lines in  $x_{1t}$  in the above asymptotic representation will cancel out, up to  $o_p(1)$ , when matrix  $R$  is applied.

Therefore, in the calculus of the difference of these rows in the formula of the test through matrix  $R$ , the corresponding difference components collapse, and the rank of matrix  $RC^{-1}R'$  is asymptotically equal to  $K_1 + G$  minus the number of collapsing components (i.e.  $K_1$ ), which yields a singular matrix of rank  $G$ . This also implies to use a generalised inverse matrix for  $RC^{-1}R'$ . This completes the proof. QED.

**Proof of Lemma 1:** Under Assumptions 1 and 2, the cross-product estimators  $\hat{Q}_z$ ,  $\hat{Q}_{zx}$ , and  $\hat{Q}_x$  are consistent almost surely due to the Kolmogorov law of large numbers. The consistency of  $\hat{Q}_{zz}$  and  $H(\hat{\Pi})$  stems from the consistency of  $\hat{\Pi}$ . Hence, it remains to show the consistency of  $\hat{\sigma}_{11} = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{1t} \hat{\epsilon}_{1t}$ ,  $\hat{\sigma}_{22} = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{2t} \hat{\epsilon}_{2t}$  and  $\hat{\sigma}_{12} = T^{-1} \sum_{t=1}^T \hat{\epsilon}_{1t} \hat{\epsilon}_{2t}$ . The detailed proof is shown only for  $\hat{\sigma}_{12}$ , since that the same kind of arguments are used for  $\hat{\sigma}_{11}$  and  $\hat{\sigma}_{22}$ .

Recalling the definition of  $\sigma_{12} = E(\epsilon_{1t} \epsilon_{2t})$ , applying directly the Kolmogorov law of large numbers, yields:  $\hat{\sigma}_{12} - \sigma_{12} \xrightarrow{p} 0$ , where  $\tilde{\sigma}_{12} = T^{-1} \sum_{t=1}^T \epsilon_{1t} \epsilon_{2t}$ . This is possible because (i)  $\epsilon_{1t} \epsilon_{2t}$  is iid from Assumption 1, and (ii)  $E(|\epsilon_{1t} \epsilon_{2t}|) < \infty$  because function  $\psi_\theta(\cdot)$  is bounded. Hence, we just need to show  $\hat{\sigma}_{12} - \tilde{\sigma}_{12} \xrightarrow{p} 0$  to prove that  $\hat{\sigma}_{12} - \sigma_{12} \xrightarrow{p} 0$ . We have:

$$|\hat{\sigma}_{12} - \tilde{\sigma}_{12}| \leq a_T + b_T + c_T,$$

where  $a_T = \theta T^{-1} \sum_{t=1}^T |1_{[\hat{v}_t \leq 0]} - 1_{[v_t \leq 0]}|$ ,  $b_T = \theta T^{-1} \sum_{t=1}^T |1_{[\hat{u}_t \leq 0]} - 1_{[u_t \leq 0]}|$



and  $c_T = T^{-1} \sum_{t=1}^T |1_{[\hat{u}_t \leq 0]} 1_{[\hat{v}_t \leq 0]} - 1_{[u_t \leq 0]} 1_{[v_t \leq 0]}|$ . The proof for both  $a_T \xrightarrow{p} 0$  and  $b_T \xrightarrow{p} 0$  can be found in the proof of Proposition 3 in Kim and Muller (2004). Hence we only need to show that  $c_T \xrightarrow{p} 0$ , which is done by majoration as follows

$$\begin{aligned}
c_T &\leq T^{-1} \sum_{t=1}^T |1_{[\hat{u}_t \leq 0]}| \times |1_{[\hat{v}_t \leq 0]} - 1_{[v_t \leq 0]}| + T^{-1} \sum_{t=1}^T |1_{[v_t \leq 0]}| \times |1_{[\hat{u}_t \leq 0]} - 1_{[u_t \leq 0]}| \\
&\leq T^{-1} \sum_{t=1}^T |1_{[\hat{v}_t \leq 0]} - 1_{[v_t \leq 0]}| + T^{-1} \sum_{t=1}^T |1_{[\hat{u}_t \leq 0]} - 1_{[u_t \leq 0]}| \\
&\leq \theta^{-1}(a_T + b_T) \xrightarrow{p} 0.
\end{aligned}$$

QED.

Table 1(a). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1), T = 200

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.04	0.07
	0.10	0.05	0.11
	0.15	0.09	0.21
	0.20	0.14	0.31
	0.25	0.19	0.44
	0.30	0.28	0.59

Table 1(b). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1), T = 300

	$\delta$	KM	Hausman
Size	0.00	0.04	0.05
Power	0.05	0.04	0.06
	0.10	0.07	0.14
	0.15	0.13	0.28
	0.20	0.21	0.42
	0.25	0.32	0.61
	0.30	0.45	0.76

Table 1(c). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1), T = 500

	$\delta$	KM	Hausman
Size	0.00	0.04	0.04
Power	0.05	0.05	0.09
	0.10	0.12	0.22
	0.15	0.22	0.42
	0.20	0.37	0.64
	0.25	0.56	0.82
	0.30	0.72	0.94

Table 2(a). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: Cauchy distribution, T = 200

	$\delta$	KM	Hausman
Size	0.00	0.02	0.05
Power	0.05	0.33	0.79
	0.10	0.74	0.75
	0.15	0.93	0.71
	0.20	0.99	0.61
	0.25	1.00	0.54
	0.30	1.00	0.49

Table 2(b). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: Cauchy distribution, T = 300

	$\delta$	KM	Hausman
Size	0.00	0.02	0.05
Power	0.05	0.47	0.87
	0.10	0.89	0.81
	0.15	0.99	0.69
	0.20	1.00	0.62
	0.25	1.00	0.56
	0.30	1.00	0.51

Table 2(c). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: Cauchy distribution, T = 500

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.67	0.90
	0.10	0.98	0.81
	0.15	1.00	0.72
	0.20	1.00	0.63
	0.25	1.00	0.57
	0.30	1.00	0.51

Table 3(a). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with a single outlier, T = 200

	$\delta$	KM	Hausman
Size	0.00	0.03	0.06
Power	0.05	0.04	0.05
	0.10	0.06	0.06
	0.15	0.08	0.08
	0.20	0.14	0.09
	0.25	0.19	0.12
	0.30	0.27	0.15

Table 3(b). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with a single outlier, T = 300

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.04	0.06
	0.10	0.09	0.08
	0.15	0.12	0.10
	0.20	0.21	0.14
	0.25	0.31	0.18
	0.30	0.45	0.25

Table 3(c). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with a single outlier, T = 500

	$\delta$	KM	Hausman
Size	0.00	0.04	0.04
Power	0.05	0.05	0.06
	0.10	0.12	0.11
	0.15	0.22	0.17
	0.20	0.37	0.26
	0.25	0.55	0.37
	0.30	0.73	0.51

Table 4(a). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 5% contamination, T = 200

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.03	0.05
	0.10	0.05	0.06
	0.15	0.06	0.07
	0.20	0.12	0.07
	0.25	0.17	0.08
	0.30	0.23	0.09

Table 4(b). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 5% contamination, T = 300

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.04	0.06
	0.10	0.07	0.06
	0.15	0.11	0.07
	0.20	0.17	0.08
	0.25	0.27	0.10
	0.30	0.40	0.11

Table 4(c). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 5% contamination, T = 500

	$\delta$	KM	Hausman
Size	0.00	0.04	0.05
Power	0.05	0.04	0.05
	0.10	0.10	0.07
	0.15	0.18	0.07
	0.20	0.32	0.10
	0.25	0.49	0.13
	0.30	0.66	0.16

Table 5(a). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 20% contamination, T = 200

	$\delta$	KM	Hausman
Size	0.00	0.02	0.05
Power	0.05	0.02	0.06
	0.10	0.03	0.06
	0.15	0.04	0.06
	0.20	0.05	0.06
	0.25	0.08	0.06
	0.30	0.11	0.07

Table 5(b). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 20% contamination, T = 300

	$\delta$	KM	Hausman
Size	0.00	0.02	0.05
Power	0.05	0.02	0.05
	0.10	0.04	0.05
	0.15	0.05	0.05
	0.20	0.09	0.07
	0.25	0.15	0.06
	0.30	0.21	0.07

Table 5(c). Rejection probabilities by the KM test and the Hausman test (with  $\theta = 0.5$ ) for the null hypothesis of no endogeneity: N(0,1) with 2% contamination, T = 500

	$\delta$	KM	Hausman
Size	0.00	0.03	0.05
Power	0.05	0.03	0.06
	0.10	0.06	0.06
	0.15	0.11	0.05
	0.20	0.17	0.07
	0.25	0.29	0.07
	0.30	0.40	0.08