On the finite-sample theory of exogeneity tests with possibly non-Gaussian errors and weak identification *

Firmin Doko Tchatoka† Jean-Marie Dufour ‡
University of Tasmania McGill University

First version: May 2007
Revised: October 2007, November 2008, December 2009, March 2010
This version: January 2012
Compiled: January 26, 2012, 9:12

*The authors thank Atsushi Inoue, Marine Carrasco, Jan Kiviet and Benoit Perron for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec), and a Killam Fellowship (Canada Council for the Arts).

† School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart TAS 7001, Tel: +613 6226 7226, Fax:+61 3 6226 7587; e-mail: Firmin.dokotchatoka@utas.edu.au. Homepage: http://www.fdokotchatoka.com

‡ William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

ABSTRACT

We investigate the finite-sample behaviour of the Durbin-Wu-Hausman (DWH) and Revankar-Hartley (RH) specification tests with or without identification. We consider two setups based on conditioning upon the fixed instruments and parametric assumptions on the distribution of the errors. Both setups are quite general and account for non-Gaussian errors. Except for a couple of Wu (1973) tests and the RH-test, finite-sample distributions are not available for the other statistics [including the most standard Hausman (1978) statistic] even when the errors are Gaussian. In this paper, we propose an analysis of the distributions of the statistics under both the null hypothesis (level) and the alternative hypothesis (power). We provide a general characterization of the distributions of the test statistics, which exhibits useful invariance properties and allows one to build exact tests even for non-Gaussian errors. Provided such finite-sample methods are used, the tests remain valid (level is controlled) whether the instruments are strong or weak. The characterization of the distributions of the statistics under the alternative hypothesis clearly exhibits the factors that determine power. We show that all tests have low power when all instruments are irrelevant (strict non-identification). But power does exist as soon as there is one strong instrument (despite the fact overall identification may fail). We present simulation evidence which confirms our finite-sample theory.

Key words: Exogeneity tests; finite-sample; weak instruments; strict exogeneity; Cholesky error family; pivotal; identification-robust; exact Monte Carlo exogeneity tests.

JEL classification: C3; C12; C15; C52.

Contents

List of Tables

List of Assumptions, Propositions and Theorems

1. Introduction

A basic problem in econometrics consists of estimating a linear relationship where the explanatory variables and the errors might be correlated. In order to detect an endogeneity problem between explanatory variables and disturbances, researchers often apply an exogeneity test, usually by resorting to instrumental variable (IV) methods. Exogeneity tests of the type proposed by Durbin (1954), Wu (1973), Hausman (1978), and Revankar and Hartley (1973) (henceforth DWH-and RH-tests) are often used to decide whether one should apply ordinary least squares (OLS) or instrumental variable methods. One key assumption of DWH and RH-tests however, is the available instruments are strong. Not much is known, at least in finite-sample, about their behaviour when identification is deficient or weak (weak instruments).

In the last two decades, literature has emerged that has raised concerns with the quality of inferences based on conventional methods, such as instrumental variables and ordinary least squares settings, when the instruments are only weakly correlated with the endogenous regressors. Many studies have shown that even ex-post conventional large-sample approximations are misleading when instruments are weak. The literature on the "weak instruments" problem is now considerable. Several authors have proposed identification-robust procedures that are applicable even when the instruments are weak[1]. However, identification-robust procedures usually do not focus on regressor exogeneity or instrument validity. Hence, there is still a reason to be concerned when testing the exogeneity or orthogonality of a regressor.

Doko and Dufour (2008) studied the impact of instrument endogeneity on Anderson and Rubin (1949, AR-test) and Kleibergen (2002, K-test). They show that both procedures are in general consistent against the presence of invalid instruments (hence invalid for the hypothesis of interest), whether the instruments are strong or weak. However, there are cases where test consistency may not hold and their use may lead to size distortions in large samples.

In this paper, our focus is not on the validity of the instruments, as done by Doko and Dufour (2008). We question whether the standard specification tests are valid in finite-sample when: (i) errors have possibly non-Gaussian distribution, and (ii) identification is weak. In the literature, except for Wu (1973, $\mathscr{T}_1$, $\mathscr{T}_2$ tests) and the Revankar and Hartley (1973, $\mathscr{RH}$-test), finite-sample distributions are not available for the other specification test statistics [including the most standard Hausman (1978) statistic] even when model errors are Gaussian and the identification is strong. This paper aims to fulfill this gap by simultaneously addressing issues related to finite-sample theory and identification.

Staiger and Stock (1997) provided a characterization of the asymptotic distribution of Hausman type-tests [namely $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{H}_3$] under the local-to-weak instruments asymptotic. They showed that when the instruments are asymptotically irrelevant, all three tests are valid (level is controlled) but inconsistent. Furthermore, their result indicates that $\mathscr{H}_1$ and $\mathscr{H}_2$ are conservative. The authors

[1] See *e.g.* Nelson and Startz (1990a, 1990b); Dufour (1997); Bekker (1994); Phillips (1989); Staiger and Stock (1997); Wang and Zivot (1998); Dufour (2003); Stock, Wright and Yogo (2002); Kleibergen (2002); Moreira (2003); Hall, Rudebusch and Wilcox (1996); Hall and Peixe (2003); Donald and Newey (2001); Dufour (2005, 2007).

observed that the concentration parameter which characterizes instrument quality depends on the sample size and concluded that size adjustment is infeasible. In this paper, we argue that this type of conclusion may go far. The local-to-weak instruments asymptotic assumes that all instruments are asymptotically weak. When the model is partially identified, the Staiger and Stock (1997) weak instruments asymptotic may lead to misleading results. This raises the following question: how do the alternative standard specification tests behave when at least one instrument is strong?

Recently, Hahn, Ham and Moon (2010) proposed a modified Hausman test which can be used for testing the validity of a subset of instruments. Their statistic is pivotal even when the instruments are weak. The problem however, is that the null hypothesis in their study tests the orthogonality of the instruments that are excluded from the structural equation. Then, the test proposed by the authors can be viewed as an alternative way to assess the overidentification restrictions hypothesis of the model [Hansen and Singleton (1982); Hansen (1982); Sargan (1983); Cragg and Donald (1993); Hansen, Heaton and Yaron (1996); Stock and Wright (2000); and Kleibergen (2005)]. So, the problem considered by Hahn et al. (2010) is fundamentally different and less complex than testing the exogeneity of an included instrument in the structural equation, as done by Durbin (1954), Wu (1973), Hausman (1978), and Revankar and Hartley (1973).

Guggenberger (2010) investigated the asymptotic size properties of a two-stage test in the linear IV model, when in the first stage a Hausman (1978) specification test is undertaken as a pretest of exogeneity of a regressor. He showed that the asymptotic size equaled 1 for empirically relevant choices of the parameter space. He then concluded that the Hausman pre-test does not have sufficient power against correlations that are local to zero when identification is weak, while the OLS-based t-statistic takes on large values for such nonzero correlations. While we do not question the basic result of Guggenberger (2010) in this paper, we observe that his framework is the Staiger and Stock (1997) weak instruments asymptotic, hence does not account for situations where at least one instrument is strong. Hence, the conclusions by Guggenberger (2010) may be misleading when identification is partial. Doko and Dufour (2011) provide a general asymptotic framework which allows one to examine the asymptotic behaviour of DWH-tests including cases where partial identification holds.

In this paper, we only focus on finite-sample and propose two setups which are then used to study the behaviour of the DWH and RH exogeneity tests. In the first setup, we assume that the structural errors are *strictly exogenous*, *i.e.* independent of the regressors and the available instruments. This setup is quite general and does not require additional assumptions on the (supposedly) endogenous regressors and the reduced-form errors. In particular, the endogenous regressors can be arbitrarily generated by any nonlinear function of the instruments and reduced-form parameters. Furthermore, the reduced-form errors may be heteroscedastic. The second setup assumes a *Cholesky invariance property* for both structural and reduced-form errors. A similar assumption in the context of multivariate linear regressions is also made in Dufour and Khalaf (2002); and Dufour, Khalaf and Beaulieu (2010).

In both setups, we propose a finite-sample analysis of the distribution of the tests under the null hypothesis (level) and the alternative hypothesis (power), with or without identification. Our analysis provides several new insights and extensions of earlier procedures. The characterization of the finite-sample distributions of the statistics, shows that all tests are typically robust to weak instruments (level is controlled), whether the errors are Gaussian or not. This result is then used to develop exact identification-robust exogeneity (MCE) tests which are valid even when conventional asymptotic theory breaks down. In particular, the MCE tests are still applicable even if the distribution of the errors does not have moments (Cauchy-type distribution, for example). Hence, size adjustment is feasible and the conclusion by Staiger and Stock (1997) may be misleading. Moreover, the characterization of the power of the tests clearly exhibits the factors that determine power. We show that all tests have no power in the extreme case where all instruments are weak [similar to Staiger and Stock (1997) and Guggenberger (2010)], but do have power as soon as we have one strong instrument. This suggests that the DWH and RH exogeneity tests can detect an exogeneity problem even if not all model parameters are identified, provided partial identification holds. We present simulation evidence which confirms our theoretical results.

The paper is organized as follows. Section 2 formulates the model studied, and Section 4 describes the statistics. Sections 5 and 6 study the finite-sample properties of the tests with (possibly) weak instruments. Section 7 presents the exact Monte Carlo exogeneity (MCE) test procedures while Section 8 presents a simulation experiment. Conclusions are drawn in Section 9 and proofs are presented in the Appendix.

2. Framework

We consider the following standard simultaneous equations model:

$$y = Y\beta + Z_1\gamma + u\,, \tag{2.1}$$

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V\,, \tag{2.2}$$

where $y \in \mathbb{R}^T$ is a vector of observations on a dependent variable, $Y \in \mathbb{R}^{T\times G}$ is a matrix of observations on (possibly) endogenous explanatory variables $(G \geq 1)$, $Z_1 \in \mathbb{R}^{T\times k_1}$ is a matrix of observations on exogenous variables included in the structural equation of interest (2.1), $Z_2 \in \mathbb{R}^{T\times k_2}$ is a matrix of observations on the exogenous variables excluded from the structural equation, $u = (u_1, \ldots, u_T)' \in \mathbb{R}^T$ and $V = [V_1, \ldots, V_T]' \in \mathbb{R}^{T\times G}$ are disturbance matrices with mean zero, $\beta \in \mathbb{R}^G$ and $\gamma \in \mathbb{R}^{k_1}$ are vectors of unknown coefficients, $\Pi_1 \in \mathbb{R}^{k_1\times G}$ and $\Pi_2 \in \mathbb{R}^{k_2\times G}$ are matrices of unknown coefficients. We suppose that the "instrument matrix"

$$Z = [Z_1 : Z_2] \in \mathbb{R}^{T\times k} \text{ has full-column rank} \tag{2.3}$$

where $k = k_1 + k_2$, and

$$T - k_1 - k_2 > G\,, \quad k_2 \geq G\,. \tag{2.4}$$

The usual necessary and sufficient condition for identification of this model is $\operatorname{rank}(\Pi_2) = G$.

The reduced form for $[y, Y]$ can be written as:

$$y = Z_1\pi_1 + Z_2\pi_2 + v,\; Y = Z_1\Pi_1 + Z_2\Pi_2 + V\,, \tag{2.5}$$

where $\pi_1 = \gamma + \Pi_1\beta,\ \pi_2 = \Pi_2\beta,$ and $v = u + V\beta = [v_1, \ldots, v_T]'$. If any restriction is imposed on γ, we see from $\pi_2 = \Pi_2\beta$ that β is identified if and only $\mathrm{rank}(\Pi_2) = G,$ which is the usual necessary and sufficient condition for identification of this model. When $\mathrm{rank}(\Pi_2) < G,\ \beta$ is not identified and the instruments Z_2 are weak.

In this paper, we study the finite-sample properties (size and power) of the standard exogeneity tests of the type proposed by Durbin (1954), Wu (1973), Hausman (1978), and Revankar and Hartley (1973) of the null hypothesis $\mathrm{H}_0 : \mathbb{E}(Y'u) = 0,$ including when identification is deficient or weak (weak instruments) and the errors $[u, V]$ may not have a Gaussian distribution.

3. Notations and definitions

Let $\hat{\beta} = (Y'M_1Y)^{-1}Y'M_1y$ and $\tilde{\beta} = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)y$ denote the ordinary least squares (OLS) estimator and two-stage least squares (2SLS) estimator of β respectively, where

$$M = M(Z) = I - Z(Z'Z)^{-1}Z', \quad M_1 = M(Z_1) = I - Z_1(Z_1'Z_1)^{-1}Z_1', \tag{3.1}$$

$$M_1 - M = M_1Z_2(Z_2'M_1Z_2)^{-1}Z_2'M_1. \tag{3.2}$$

Let $\hat{V} = MY,\ X = [X_1 : \hat{V}],\ X_1 = [Y : Z_1],\ \hat{X} = [\hat{X}_1 : \hat{V}],\ \hat{X}_1 = [\hat{Y} : Z_1],\ \bar{X} = [X_1 : Z_2] = [Y : Z],$ and consider the following regression of u on the columns of V :

$$u = Va + \varepsilon\,,$$

where a is a $G \times 1$ vector of unknown coefficients, and ε is independent of V with mean zero and variance σ_ε^2. Define $\theta = (\beta', \gamma', a')',\ \theta_* = (\beta', \gamma', b')',\ \bar{\theta} = (b', \bar{\gamma}', \bar{a}')',$ where $b = \beta + a,$ $\bar{\gamma} = \gamma - \Pi_1 a,\ \bar{a} = -\Pi_2 a.$ We then observe that $Y = \hat{Y} + \hat{V},$ where $\hat{Y} = (I - M)Y = P_Z Y,$ and $\beta = b$ as soon as $a = 0.$ From the above definitions and notations, the structural equation (2.1) can be written in the following three different ways:

$$y = Y\beta + Z_1\gamma + \hat{V}a + e_* = X\theta + e_*\,, \tag{3.3}$$

$$y = \hat{Y}\beta + Z_1\gamma + \hat{V}b + e_* = \hat{X}\theta_* + e_*\,, \tag{3.4}$$

$$y = Yb + Z_1\bar{\gamma} + Z_2\bar{a} + \varepsilon = \bar{X}\bar{\theta} + \varepsilon\,, \tag{3.5}$$

where $e_* = P_Z Va + \varepsilon.$ Equations (**??**)-(3.5) clearly illustrate that the endogeneity of Y may be viewed as a problem of omitted variables [see Dufour (1987)].

Let us denote by $\hat{\theta}$: the OLS estimate of θ in (3.3), $\hat{\theta}_0$: the restricted OLS estimate of θ under $a = 0,$ in (3.3), $\hat{\theta}_*$: the OLS estimate of θ_* in (3.4), $\hat{\theta}_{*0}$: the restricted OLS estimate of θ_* under $\beta = b$ in (3.4), $\hat{\theta}_*^0$: the restricted OLS estimate of θ_* under $b = 0$ in (3.4) or $\beta = -a$ in (3.3),$\hat{\bar{\theta}}$:

the OLS estimate of $\bar{\theta}$ in (3.5), $\hat{\bar{\theta}}_0$: the restricted OLS estimate of $\bar{\theta}$ under $\bar{a}=0$, and define the following sum squared errors:

$$\begin{aligned} S(\omega) &= (y-X\omega)'(y-X\omega), S_*(\omega)=(y-\hat{X}\omega)'(y-\hat{X}\omega), \\ \bar{S}(\omega) &= (y-\bar{X}\omega)'(y-\bar{X}\omega), \quad \forall\,\omega\in\mathbb{R}^{k_1+2G}. \end{aligned} \tag{3.6}$$

Let

$$\tilde{\Sigma}_1 = \tilde{\sigma}_1^2\hat{\Delta}\,, \quad \tilde{\Sigma}_2=\tilde{\sigma}_2^2\hat{\Delta}\,, \quad \tilde{\Sigma}_3=\tilde{\sigma}^2\hat{\Delta}\,, \quad \tilde{\Sigma}_4=\hat{\sigma}^2\hat{\Delta}\,, \tag{3.7}$$

$$\hat{\Sigma}_1 = \tilde{\sigma}^2\hat{\Omega}_{IV}^{-1}-\hat{\sigma}^2\hat{\Omega}_{LS}^{-1}\,, \quad \hat{\Sigma}_2=\tilde{\sigma}^2\hat{\Delta}\,, \quad \hat{\Sigma}_3=\hat{\sigma}^2\hat{\Delta}\,, \tag{3.8}$$

$$\hat{\Sigma}_R = \frac{1}{\hat{\sigma}_R^2}D_1Z_2(Z_2'D_1Z_2)^{-1}Z_2'D_1, D_1=\frac{1}{T}M_1M_{M_1Y}\,, \tag{3.9}$$

$$\hat{\Omega}_{IV} = \frac{1}{T}Y'(M_1-M)Y\,, \quad \hat{\Omega}_{LS}=\frac{1}{T}Y'M_1Y, \hat{\Delta}=\hat{\Omega}_{IV}^{-1}-\hat{\Omega}_{LS}^{-1}, \tag{3.10}$$

where $\hat{\sigma}^2=(y-Y\hat{\beta})'M_1(y-Y\hat{\beta})/T$ is the OLS-based estimator of σ_u^2, $\tilde{\sigma}^2=(y-Y\tilde{\beta})'M_1(y-Y\tilde{\beta})/T$ is the usual 2SLS-based estimator of σ_u^2 (both without correction for degrees of freedom), while $\tilde{\sigma}_1^2=(y-Y\tilde{\beta})'(M_1-M)(y-Y\tilde{\beta})/T=\tilde{\sigma}^2-\tilde{\sigma}_e^2$, $\tilde{\sigma}_2^2=\hat{\sigma}^2-(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta})=\hat{\sigma}^2-\tilde{\sigma}^2(\tilde{\beta}-\hat{\beta})'\hat{\Sigma}_2^{-1}(\tilde{\beta}-\hat{\beta})$, $\tilde{\sigma}_e^2=(y-Y\tilde{\beta})'M(y-Y\tilde{\beta})/T$, and $\hat{\sigma}_R^2=yM_{\bar{X}}y'/T$ may be interpreted as alternative IV-based scaling factors, $\kappa_1=(k_2-G)/G$, $\kappa_2=(T-k_1-2G)/G$, $\kappa_3=\kappa_4=T-k_1-G$, and $\kappa_R=(T-k_1-k_2-G)/k_2$. From (3.6) and (3.7)-(3.10), we can see that

$$S(\hat{\theta}) = S_*(\hat{\theta}_*),\, S(\hat{\theta}_0)=S_*(\hat{\theta}_{*0}), S(\hat{\theta})=T\tilde{\sigma}_2^2,\, S(\hat{\theta}_0)=T\hat{\sigma}^2,\, S_*(\hat{\theta}_*^0)=T\tilde{\sigma}^2. \tag{3.11}$$

Throughout the paper, we also adopt the following notations:

$$\begin{aligned}
C_0 &= (\bar{A}_1-A_1)'\hat{\Delta}^{-1}(\bar{A}_1-A_1),\, \bar{A}_1=[Y'(M_1-M)Y]^{-1}Y'(M_1-M), && (3.12)\\
A_1 &= (Y'M_1Y)^{-1}Y'M_1,\, \bar{D}_1=\frac{1}{T}M_1M_{(M_1-M)Y},\, D_1=\frac{1}{T}M_1M_{M_1Y}, && (3.13)\\
\Sigma_1 &= (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\hat{\Omega}_{IV}^{-1}-(Va+\varepsilon)'D_1(Va+\varepsilon)\hat{\Omega}_{LS}^{-1}\,, && (3.14)\\
\Omega_{IV} &\equiv \Omega_{IV}(\mu_2,\bar{V})=(\mu_2+\bar{V})'(M_1-M)(\mu_2+\bar{V})\,, && (3.15)\\
\Omega_{LS} &\equiv \Omega_{LS}(\mu_2,\bar{V})=(\mu_2+\bar{V})'M_1(\mu_2+\bar{V})\,, && (3.16)\\
\omega_{IV}^2 &\equiv \omega_{IV}(\mu_1,\mu_2,\bar{V},\bar{v})^2=(\mu_1+\bar{v})'D_*'D_*(\mu_1+\bar{v})\,, && (3.17)\\
\omega_{LS}^2 &\equiv \omega_{LS}(\mu_1,\mu_2,\bar{V},\bar{v})^2=(\mu_1+\bar{v})'C_*(\mu_1+\bar{v})\,, && (3.18)\\
C_* &= M_1-M_1(\mu_2+\bar{V})\Omega_{LS}(\mu_2,\bar{V})^{-1}(\mu_2+\bar{V})'M_1 && (3.19)\\
D_* &= M_1-M_1(\mu_2+\bar{V})\Omega_{IV}(\mu_2,\bar{V})^{-1}(\mu_2+\bar{V})'(M_1-M)\,, && (3.20)\\
\omega_1^2 &\equiv \omega_1(\mu_1,\mu_2,\bar{V},\bar{v})^2=(\mu_1+\bar{v})'E(\mu_1+\bar{v})\,, && (3.21)\\
\omega_2^2 &\equiv \omega_2(\mu_1,\mu_2,\bar{V},\bar{v})^2=(\mu_1+\bar{v})'[C_*-C'\Delta^{-1}C](\mu_1+\bar{v})\,, && (3.22)\\
\omega_R^2 &\equiv \omega_R(\mu_1,\mu_2,\bar{V},\bar{v})^2=(\mu_1+\bar{v})'[D_1-P_{D_1Z_2}](\mu_1+\bar{v})\,, && (3.23)\\
C &= \Omega_{IV}(\mu_2,\bar{V})^{-1}(\mu_2+\bar{V})'(M_1-M)-\Omega_{LS}(\mu_2,\bar{V})^{-1}(\mu_2+\bar{V})'M_1, && (3.24)
\end{aligned}$$

$$
\begin{aligned}
E &= (M_1 - M)\left[I - (\mu_2 + \bar{V})\Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'\right](M_1 - M), && (3.25)\\
\omega_3^2 &\equiv \omega_3(\mu_1, \mu_2, \bar{V}, \bar{v})^2 = \omega_{IV}^2,\ \omega_4^2 \equiv \omega_4(\mu_1, \mu_2, \bar{V}, \bar{v})^2 = \omega_{LS}^2, && (3.26)\\
\Gamma_1(\mu_1, \mu_2, \bar{V}, \bar{v}) &= C'[\omega_{IV}^2\Omega_{IV}^{-1} - \omega_{LS}^2\Omega_{LS}^{-1}]^{-1}C, \Gamma_2(\mu_1, \mu_2, \bar{V}, \bar{v}) = \frac{1}{\omega_{IV}^2}C'\Delta^{-1}C, && (3.27)\\
\Gamma_3(\mu_1, \mu_2, \bar{V}, \bar{v}) &= \frac{1}{\omega_{LS}^2}C'\Delta^{-1}C,\ \bar{\Gamma}_l(\mu_1, \mu_2, \bar{V}, \bar{v}) = \frac{1}{\omega_l^2}C'\Delta^{-1}C, \quad l = 1, 2, 3, 4, && (3.28)\\
\Gamma_R(\mu_1, \mu_2, \bar{V}, \bar{v}) &= \frac{1}{\omega_R^2}P_{D_1Z_2}, && (3.29)
\end{aligned}
$$

where for any matrix B, $P_B = B(B'B)^{-1}B'$ is the projection matrix on the space spanned by the columns of B, and $M_B = I - P_B$.

Finally, let $C_\pi = \Pi_2'Z_2'M_1Z_2\Pi_2$ denotes the concentration factor[2]. We then have $M_1Z_2\Pi_2 a = 0$ if and only if $C_\pi a = 0$, *i.e.* $a = (I_G - C_\pi^- C_\pi)a^*$, where C_π^- is any generalized inverse of C_π, and a^* is an arbitrary $G \times 1$ vector [see Rao and Mitra (1971, Theorem 2.3.1)]. Let

$$
\mathcal{N}(C_\pi) = \{\varpi \in \mathbb{R}^G : C_\pi \varpi = 0\}, \tag{3.30}
$$

denotes the null set of the linear map on $\mathbb{R}^G$ characterized by the matrix C_π. Observe that when $Z_2'M_1Z_2$ has a full column rank k_2, $\mathcal{N}(C_\pi) = \{\varpi \in \mathbb{R}^G : \Pi_2\varpi = 0\}$. Hence, provided identification holds, $\mathcal{N}(C_\pi) = \{0\}$. However, when identification is weak or deficient, there exist $\varpi_0 \neq 0$ such that $\varpi_0 \in \mathcal{N}(C_\pi)$.

Section 4 presents the DWH and RH test statistics.

4. Exogeneity test statistics

We consider Durbin-Wu-Hausman test statistics, namely three versions of the Hausman-type statistics $[\mathcal{H}_i, i = 1, 2, 3]$, the four statistics proposed by Wu (1973) $[\mathcal{T}_l, l = 1, 2, 3, 4]$ and the test statistic proposed by Revankar and Hartley (1973, RH). First, we propose a unified presentation of these statistics that shows the link between Hausman-and Wu-type tests. Second, we provide an alternative derivation of all test statistics (including RH test statistic) from the residuals of the regression of the unconstrained and constrained models.

4.1. Unified presentation

This subsection proposes a unified presentation of the DWH and RH test statistics. The proof of this unified representation is attached in Appendix A.1. The four statistics proposed by Wu (1973) can all be written in the form

$$
\mathcal{T}_l = \kappa_l(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta} - \hat{\beta}), \quad l = 1, 2, 3, 4. \tag{4.1}
$$

[2] If the errors V have a definite positive covariance matrix Σ_V, then $\Sigma_V^{-\frac{1}{2}}C_\pi\Sigma_V^{-\frac{1}{2}}$ is often referred to as the concentration matrix. Hence, we referred to C_π as the concentration factor.

The three versions of Hausman-type statistics are defined as

$$\mathcal{H}_i = T(\tilde{\beta}-\hat{\beta})'\hat{\Sigma}_i^{-1}(\tilde{\beta}-\hat{\beta}), \quad i=1,2,3. \tag{4.2}$$

And the Revankar and Hartley (1973, RH) statistic is given by:

$$\mathcal{RH} = \kappa_R y'\hat{\Sigma}_R y. \tag{4.3}$$

The corresponding tests reject H_0 when the test statistic is “large”. Unlike $\mathcal{RH}$, $\mathcal{H}_i, i=1,2,3,$ and $\mathcal{T}_l, l=1,2,3,4,$ compare OLS to 2SLS estimators of β. They only differ through the use of different “covariance matrices”. $\mathcal{H}_1$ uses two different estimators of σ_u^2, while the others resort to a single scaling factor (or estimator of σ_u^2). The expressions of the $\mathcal{T}_l, l=1,2,3,4,$ in (4.1) are much more interpretable than those in Wu (1973). The link between Wu (1973) notations and ours is established in Appendix A.1. We use the above notations to better see the relation between Hausman-type tests and Wu-type tests. In particular, it is easy to see that $\tilde{\Sigma}_3=\hat{\Sigma}_2$ and $\tilde{\Sigma}_4=\hat{\Sigma}_3$, so $\mathcal{T}_3=(\kappa_3/T)\mathcal{H}_2$ and $\mathcal{T}_4=(\kappa_4/T)\mathcal{H}_3$.

Finite-sample distributions are available for $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{RH}$ when the errors are Gaussian. More precisely, if $u\sim N[0,\sigma^2 I_T]$ and Z is independent of u, then:

$$\mathcal{T}_1\sim F(G,k_2-G), \quad \mathcal{T}_2\sim F(G,T-k_1-2G), \quad \mathcal{RH}\sim F(k_2,T-k_1-k_2-G) \tag{4.4}$$

under the null hypothesis of exogeneity. If furthermore, $\text{rank}(\Pi_2)=G$ and the sample size is large, under the exogeneity of Y, we have (with standard regularity conditions):

$$\mathcal{H}_i \xrightarrow{L} \chi^2(G), i=1,2,3; \mathcal{T}_l \xrightarrow{L} \chi^2(G), l=3,4. \tag{4.5}$$

However, even when identification is strong and the errors Gaussian, the finite-sample distributions of $\mathcal{H}_i$, $i=1,2,3$ and $\mathcal{T}_l$, $l=3,4$ are not established in the literature. This underscores the importance of this study.

4.2. Regression interpretation

We now give the regression interpretation of the above statistics. From Section 3, except for $\mathcal{H}_1$, $\mathcal{H}_i, i=2,3,$ $\mathcal{T}_l, l=1,2,3,4$ and $\mathcal{RH}$ can be expressed as [see Appendix A.2 for further details]:

$$\mathcal{H}_2 = T[S(\hat{\theta}_0)-S(\hat{\theta})]/S_*(\hat{\theta}_*^0), \mathcal{H}_3 = T[S(\hat{\theta}_0)-S(\hat{\theta})]/S(\hat{\theta}_0), \tag{4.6}$$
$$\mathcal{T}_1 = \kappa_1[S(\hat{\theta}_0)-S(\hat{\theta})]/[S_*(\hat{\theta}_*^0)-S_e(\hat{\theta})], \mathcal{T}_2=\kappa_2[S(\hat{\theta}_0)-S(\hat{\theta})]/S(\hat{\theta}), \tag{4.7}$$
$$\mathcal{T}_3 = \kappa_3[S(\hat{\theta}_0)-S(\hat{\theta})]/S_*(\hat{\theta}_*^0), \mathcal{T}_4=\kappa_4[S(\hat{\theta}_0)-S(\hat{\theta})]/S(\hat{\theta}_0), \tag{4.8}$$
$$\mathcal{RH} = \kappa_R[\bar{S}(\bar{\hat{\theta}}_0)-\bar{S}(\bar{\hat{\theta}})]/\bar{S}(\bar{\hat{\theta}}_0), \tag{4.9}$$

where $S_e(\hat{\theta})=T\tilde{\sigma}_e^2$. Equations (4.6) - (4.9) are the regression formulation of the DWH and RH statistics. It interesting to observe that DWH statistics test the null hypothesis $H_0: a=0,$ while RH tests $\mathrm{H}_0^*: \bar{a} = -\Pi_2 a = 0$. If $\mathrm{rank}(\Pi_2) = G$, $a = 0$ if and only if $\bar{a} = 0$. However, if $\mathrm{rank}(\Pi_2) < G$, $\bar{a} = 0$ does not entail $a = 0$. So, $\mathrm{H}_0 \subseteq \mathrm{H}_0^*$ but the inverse may not hold.

Our analysis of the distribution of the statistics under the null hypothesis (level) and the alternative hypothesis (power), considers two setups. The first setup is *the strict exogeneity*, *i.e.* the structural error u is independent of all regressors. The second setup is *the Cholesky error family.* This setup assumes that the reduced-form errors belong to Cholesky families.

5. Strict exogeneity

In this section, we consider the problem of testing the strict exogeneity of Y, *i.e.* the problem:

$$\mathrm{H}_0 \quad : \quad u \text{ is independent of } [Y, Z] \tag{5.1}$$

vs

$$H_1 \quad : \quad u = Va + \varepsilon\,, \tag{5.2}$$

where a is a $G \times 1$ vector of unknown coefficients, ε is independent of V with mean zero and variance σ_ε^2. It is important to observe that equation (5.2) does not impose restrictions on the structure of the errors u and V. This equation is interpreted as the projection of u in the columns of V and holds for any homoscedastic disturbances u and V with mean zero. Thus, the hypothesis H_0 can be expressed as

$$\mathrm{H}_0 \quad : \quad a = 0\,. \tag{5.3}$$

Note that (5.1) - (5.2) do not require any assumption concerning the functional form of Y. So, we could assume that Y obeys a general model of the form:

$$Y \quad = \quad g(Z_1, Z_2, V, \Pi)\,, \tag{5.4}$$

where $g(.)$ is a possibly unspecified non-linear function, Π is an unknown parameter matrix and V follows an arbitrary distribution. This setup is quite wide and does allow one to study several situations where neither V nor u follow a Gaussian distribution. This is particularly important in financial models with fat-tailed error distributions, such as the Student-t. Furthermore, the errors u and V may not have moments (Cauchy distribution for example).

Section 5.1 studies the distributions of the statistics under the null hypothesis (level).

5.1. Pivotality under strict exogeneity

We first characterize the finite-sample distributions of the statistics under H_0, including when identification is weak and the errors are possibly non-Gaussian. Theorem **5.1** establishes the pivotality of all statistics.

Theorem 5.1 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Suppose the assumptions* (2.1), (2.3) - (2.4) *hold. Under* H_0, *the conditional distributions given* $[Y : Z]$ *of all statistics defined by* (4.1) - (4.3) *depend only on the distribution of* u/σ_u *irrespective of whether the instruments are strong or weak.*

The results of Theorem **5.1** indicate that if the conditional distribution of $(u/\sigma_u)|Y,Z$ does not involve any nuisance parameter, then all exogeneity tests are typically robust to weak instruments (level is controlled) whether the instruments are strong or weak. More interestingly, this holds even if $(u/\sigma_u)|Y,Z$ do not follow a Gaussian distribution. As a result, exact identification-robust procedures can be developed from the standard specification test statistics even when the errors have a non-Gaussian distribution (see Section 7). This is particularly important in financial models with fat-tailed error distributions, such as the Student-t or in models where the errors may not have any moment (Cauchy-type errors, for example). Furthermore, the exact procedures proposed in Section 7 do not require any assumption on the distribution of V and the functional form of Y. More generally, one could assume that Y obeys a general non-linear model as defined in (5.4) and that $V_1, \ldots, V_T$ are heteroscedastic.

Section 5.2 characterizes the power of the tests.

5.2. Power under strict exogeneity

We characterize the distributions of the tests under the general hypothesis (5.2). As before, we cover both weak and strong identification setups. Theorem **5.2** presents the results.

Theorem 5.2 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Let the assumptions* (2.1) - (2.4) *hold. If furthermore* H_1 *in* (5.2) *is satisfied, then we can write*

$$\mathscr{H}_1 = T(Va+\varepsilon)'(\bar{A}_1 - A_1)'\Sigma_1^{-1}(\bar{A}_1 - A_1)(Va+\varepsilon)\,, \tag{5.5}$$
$$\mathscr{H}_2 = T(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\,, \tag{5.6}$$
$$\mathscr{H}_3 = T(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'D_1(Va+\varepsilon)\,, \tag{5.7}$$
$$\mathscr{T}_1 = \kappa_1(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'(\bar{D}_1 - D_1)(Va+\varepsilon)\,, \tag{5.8}$$
$$\mathscr{T}_2 = \kappa_2(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'(D_1 - C_0)(Va+\varepsilon)\,, \tag{5.9}$$
$$\mathscr{T}_3 = \kappa_3(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\,, \tag{5.10}$$
$$\mathscr{T}_4 = \kappa_4(Va+\varepsilon)'C_0(Va+\varepsilon)/(Va+\varepsilon)'D_1(Va+\varepsilon)\,, \tag{5.11}$$
$$\mathscr{RH} = \kappa_R(Va+\varepsilon)'P_{D_1Z_2}(Va+\varepsilon)/(Va+\varepsilon)'(D_1 - P_{D_1Z_2})(Va+\varepsilon)\,, \tag{5.12}$$

where Σ_1, C_0, A_1, $\bar{D}_1$, D_1, $\hat{\Omega}_{IV}$, $\hat{\Omega}_{LS}$, $\hat{\Delta}$ κ_R, *and* κ_l, $l = 1, 2, 3, 4$, *are defined in Section* 3.

We note first that Theorem **5.2** follows from algebraic arguments only. So, $[Y : Z]$ can be random in any arbitrary way. Second, given $[Y : Z]$, the distributions of the statistics only depend on the endogeneity a. We Can then observe that the above characterization clearly exhibits $(\bar{A}_1 - A_1)Va$,

C_0Va, D_1Va, $\bar{D}_1Va$, $P_{D_1Z_2}Va$ as the factors that determine power. As a result, Corollary **5.3** examine the case where all exogeneity tests do not have power.

Corollary 5.3 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem* **5.2**, *all exogeneity tests do not have power if and on if* $a \in \mathscr{N}(C_\pi)$. *More precisely, the following equalities:*

$$\begin{aligned} \mathscr{H}_1 &= T\varepsilon'(\bar{A}_1 - A_1)'\Sigma_{1*}^{-1}(\bar{A}_1 - A_1)\varepsilon\,, && (5.13)\\ \mathscr{H}_2 &= T\varepsilon' C_0\varepsilon/\varepsilon'\bar{D}_1\varepsilon,\ \mathscr{H}_3 = T\varepsilon' C_0\varepsilon/\varepsilon' D_1\varepsilon\,, && (5.14)\\ \mathscr{T}_1 &= \kappa_1\varepsilon' C_0\varepsilon/\varepsilon'(\bar{D}_1 - D_1)\varepsilon,\ \mathscr{T}_2 = \kappa_2\varepsilon' C_0\varepsilon/\varepsilon'(D_1 - C_0)\varepsilon\,, && (5.15)\\ \mathscr{T}_3 &= \kappa_3\varepsilon' C_0\varepsilon/\varepsilon'\bar{D}_1\varepsilon,\ \mathscr{T}_4 = \kappa_4\varepsilon' C_0\varepsilon/\varepsilon' D_1\varepsilon\,, && (5.16)\\ \mathscr{R}\mathscr{H} &= \kappa_R\varepsilon' P_{D_1Z_2}\varepsilon/\varepsilon'(D_1 - P_{D_1Z_2})\varepsilon && (5.17) \end{aligned}$$

hold with probability 1 if and only if $a \in \mathscr{N}(C_\pi)$, *where* $\Sigma_{1*} = \varepsilon'\bar{D}_1\varepsilon\hat{\Omega}_{IV}^{-1} - \varepsilon' D_1\varepsilon\hat{\Omega}_{LS}^{-1}$.

When $a \in \mathscr{N}(C_\pi)$, the conditional distributions of the statistics, given $[Y : Z]$, are the same under the null hypothesis and the alternative hypothesis. Therefore, their unconditional distributions are also the same under the null and the alternative hypotheses. This entails that the power of the tests can not exceed the nominal level. This condition is satisfied when $\Pi_2 = 0$ (irrelevant instruments), and all exogeneity tests have no power against complete non identification of model parameters.

We now analyze the properties of the tests when model errors belong to Cholesky families.

6. Cholesky error families

Let

$$U = [u, V] = [U_1, \ldots, U_T]'\,, \tag{6.18}$$

$$W = [v, V] = [u + V\beta, V] = [W_1, W_2, \ldots, W_T]'\,. \tag{6.19}$$

We assume that the vectors $U_t = [u_t, V_t']'$, $t = 1, \ldots, T$, have the same nonsingular covariance matrix:

$$E[U_tU_t'] = \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0\,, \quad t = 1, \ldots, T, \tag{6.20}$$

where Σ_V has dimension G. Then the covariance matrix of the reduced-form disturbances $W_t = [v_t, V_t']'$ also have the same covariance matrix, which takes the form:

$$\Omega = \begin{bmatrix} \sigma_u^2 + \beta'\Sigma_V\beta + 2\beta'\delta & \beta'\Sigma_V + \delta' \\ \Sigma_V\beta + \delta & \Sigma_V \end{bmatrix} \tag{6.21}$$

where Ω is positive definite. In this framework, the exogeneity hypothesis can be expressed as

$$\mathrm{H}_0 : \delta = 0\,. \tag{6.22}$$

Suppose that equation (5.2) holds, we can see from (6.20) that

$$\delta = \Sigma_V a, \quad \sigma_u^2 = \sigma_\varepsilon^2 + a'\Sigma_V a = \sigma_\varepsilon^2 + \delta'\Sigma_V^{-1}\delta\,. \tag{6.23}$$

So, the null hypothesis in (6.22) can be expressed as

$$H_a : a = 0\,. \tag{6.24}$$

Assume that

$$W_t = J\bar{W}_t\,, t = 1, \ldots, T\,, \tag{6.25}$$

where the vector $W_{(T)} = vec(\bar{W}_1, \ldots, \bar{W}_T)$ has a known distribution $F_{\bar{W}}$ and $J \in R^{(G+1)\times(G+1)}$ is an unknown upper triangular nonsingular matrix [for a similar assumption in the context of multivariate linear regressions, see Dufour and Khalaf (2002) and Dufour et al. (2010)]. When the errors W_t obey (6.25), we say that W_t belongs to the Cholesky error family.

If the covariance matrix of $\bar{W}_t$ is an identity matrix I_{G+1}, the covariance matrix of W_t is

$$\Omega = E[W_t W_t'] = JJ'. \tag{6.26}$$

In particular, these conditions are satisfied when

$$\bar{W}_t \overset{i.i.d.}{\sim} N[0, I_{G+1}]\,, t = 1, \ldots, T\,. \tag{6.27}$$

Since the J matrix is upper triangular, its inverse J^{-1} is also upper triangular. Let

$$P = (J^{-1})'. \tag{6.28}$$

Clearly, P is a $(G+1)\times(G+1)$ lower triangular matrix and it allows one to orthogonalize JJ' :

$$P'JJ'P = I_{G+1}\,, \quad (JJ')^{-1} = PP'\,. \tag{6.29}$$

In (6.29), P' can be interpreted as the Cholesky factor of Ω^{-1}, so P is the unique lower triangular matrix that satisfies equation (6.29); see Harville (1997, Section 14.5, Theorem 14.5.11). We will find useful to consider the following partition of P :

$$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \tag{6.30}$$

where $P_{11} \neq 0$ is a scalar and P_{22} is a nonsingular $G \times G$ matrix. In particular, if (6.26) holds, we see [using (6.21)] that an appropriate P matrix is obtained by taking:

$$P_{11} = (\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{-1/2} = \sigma_\varepsilon\,, \quad P_{22}'\Sigma_V P_{22} = I_G\,, \tag{6.31}$$

$$P_{21} = -(\beta + \Sigma_V^{-1}\delta)(\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{-1/2} = -(\beta + a)\sigma_\varepsilon^{-1}. \tag{6.32}$$

Further this choice is unique. From (6.32), P_{22} only depends on Σ_V and $P_{11}\beta + P_{21} = -(\Sigma_V^{-1}\delta)\sigma_\varepsilon^{-1} = -a\sigma_\varepsilon^{-1}$. In particular, if $\delta = 0$, we have $P_{11} = 1/\sigma_u$, $P_{21} = -\beta/\sigma_u$ and $P_{11}\beta + P_{21} = 0$.

If we postmultiply $[y, Y]$ by P, we obtain from (2.5):

$$[\bar{y}, \bar{Y}] = [y, Y]P = [yP_{11} + YP_{21}, YP_{22}] = [Z_1, Z_2] \begin{bmatrix} \gamma + \Pi_1\beta & \Pi_1 \\ \Pi_2\beta & \Pi_2 \end{bmatrix} P + \bar{W} \tag{6.33}$$

where

$$\bar{W} = UP = [\bar{v}, \bar{V}] = [\bar{W}_1, \ldots, \bar{W}_T]', \; \bar{W}_t = [\bar{v}_t, \bar{V}_t']', \tag{6.34}$$

$$\bar{v} = vP_{11} + VP_{21} = [\bar{v}_1, \ldots, \bar{v}_T]', \quad \bar{V} = VP_{22} = [\bar{V}_1, \ldots, \bar{V}_T]'. \tag{6.35}$$

Then, we can rewrite (6.33) as

$$\bar{y} = Z_1(\gamma P_{11} + \Pi_1\zeta) + Z_2\Pi_2\zeta + \bar{v}, \tag{6.36}$$

$$\bar{Y} = Z_1\Pi_1P_{22} + Z_2\Pi_2P_{22} + \bar{V}, \tag{6.37}$$

where

$$\zeta = \beta P_{11} + P_{21} = -(\Sigma_V^{-1}\delta)/(\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{1/2} = -a\sigma_\varepsilon^{-1}. \tag{6.38}$$

Since $MZ = 0$, we have

$$M\bar{y} = M\bar{v}, M\bar{Y} = M\bar{V}, \tag{6.39}$$

$$M_1\bar{y} = M_1(\mu_1 + \bar{v}), M_1\bar{Y} = M_1(\mu_2 + \bar{V}). \tag{6.40}$$

where

$$\mu_1 = M_1Z_2\Pi_2\zeta = -\sigma_\varepsilon^{-1}M_1Z_2\Pi_2 a,$$

$$\mu_2 = M_1Z_2\Pi_2P_{22}. \tag{6.41}$$

Clearly, μ_2 does not depend on the endogeneity parameter $a = \Sigma_V^{-1}\delta$. Furthermore, $\zeta = 0 \Leftrightarrow \delta = a = 0$ and $\mu_1 = 0$. In particular, this condition holds under H_0 $(\delta = a = 0)$. If $\Pi_2 = 0$ (complete non-identification of the model parameters), we have $\mu_1 = 0$ and $\mu_2 = 0$, irrespective of the value of δ. In this case,

$$M\bar{y} = M\bar{v}, M\bar{Y} = M\bar{V}, M_1\bar{y} = M_1\bar{v}, M_1\bar{Y} = M_1\bar{V}. \tag{6.42}$$

We can now prove the following Cholesky invariance property of all test statistics.

Lemma 6.1 CHOLESKY INVARIANCE OF EXOGENEITY TESTS. *Let*

$$R = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \tag{6.43}$$

be a lower triangular matrix such that $R_{11} \neq 0$ *is a scalar and* R_{22} *is a nonsingular* $G \times G$ *matrix. If we replace* y *and* Y *by* $y_* = yR_{11} + YR_{21}$ *and* $Y_* = YR_{22}$ *in* (4.1) - (3.10)*, then the statistics* H_i $(i = 1, 2, 3)$, T_l $(l = 1, 2, 3, 4)$ *and RH do not change.*

The above invariance holds irrespective of the choice of lower triangular matrix R. In particular, one can choose $R = P$ as defined in (6.28). We can now prove the following general theorem on the distributions of the test statistics.

Theorem 6.2 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions* (2.1) - (2.4) *and assumption* (6.25)*, the statistics defined in* (4.1) - (3.10) *have the following representations:*

$$\mathscr{H}_i = T[\mu_1 + \bar{v}]'\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}], \quad i = 1, 2, 3,$$

$$\mathscr{T}_l = \kappa_l[\mu_1 + \bar{v}]'\bar{\Gamma}_l(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}], \quad l = 1, 2, 3, 4,$$

$$\mathscr{R}\mathscr{H} = \kappa_R[\mu_1 + \bar{v}]'\Gamma_R(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}],$$

where $[\bar{v}, \bar{V}]$, μ_1, μ_2 *are defined in* (6.34) *and* (6.41), Γ_i, $\bar{\Gamma}_l$, *and* Γ_R *are defined in Section* 3.

The above theorem entails that the distributions of the statistics do not depend on either β or γ. Observe that Theorem **6.2** follows from algebraic arguments only, so $[Y, Z]$ and $[\bar{v}, \bar{V}]$ can be random in an arbitrary way. If the distributions of Z and $[\bar{v}, \bar{V}]$ do not depend on other model parameters, the theorem entails that the distributions of the statistics depend on model parameters only through μ_1 and μ_2. Since μ_2 does not involve δ, μ_1 is the only factor that determines power. If $\mu_1 \neq 0$, the tests have power. This may be the case when at least one instrument is strong (partial identification of model parameters). However, we can observe that when $M_1Z_2\Pi_2 a = 0$, $\mu_1 = 0$ and exogeneity tests have no power. We now provide a formal characterization of the set of parameters in which exogeneity tests have no power.

Corollary **6.3** characterizes the power of the tests when $a \in \mathscr{N}(C_\pi)$.

Corollary 6.3 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem* **6**.**2**, *if* $a \in \mathscr{N}(C_\pi)$, *we have* $\mu_1 = 0$ *and the statistics defined in* (4.1) - (3.10) *have the following representations:*

$$\begin{aligned} \mathscr{H}_i &= T\bar{v}'\Gamma_i(\mu_2, \bar{v}, \bar{V})\bar{v}, \, i = 1, 2, 3; \, \mathscr{T}_l = \kappa_l \bar{v}'\bar{\Gamma}_l(\mu_2, \bar{v}, \bar{V})\bar{v}, \, l = 1, 2, 3, 4, \\ \mathscr{R}\mathscr{H} &= \kappa_R \bar{v}'\Gamma_R(\mu_2, \bar{v}, \bar{V})\bar{v} \end{aligned}$$

irrespective of whether the instruments are weak or strong, where $\Gamma_i(\mu_2, \bar{v}, \bar{V}) \equiv \Gamma_i(0, \mu_2, \bar{v}, \bar{V})$, $\bar{\Gamma}_l(\mu_2, \bar{v}, \bar{V}) \equiv \Gamma_l(0, \mu_2, \bar{v}, \bar{V})$, $\Gamma_R(\mu_2, \bar{v}, \bar{V}) = \Gamma_R(0, \mu_2, \bar{v}, \bar{V})$, $\zeta = -(\Sigma_V^{-1}\delta)/(\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{1/2}$, Γ_i, $\bar{\Gamma}_l$, *and* Γ_R *are defined in Section* 3.

First, note that when $a \in \mathscr{N}(C_\pi)$, i.e. when $M_1 Z_2 \Pi_2 a = 0$, the conditional distributions, given Z and $\bar{V}$ of the exogeneity tests, only depend on μ_2 irrespective of the quality of the instruments. In particular, this condition is satisfied when $\Pi_2 = 0$ (complete non-identification of the model parameters) or $\delta = a = 0$ (under the null hypothesis). Since μ_2 does not depend on δ or a, all exogeneity test statistics have the same distribution under both the null hypothesis $(\delta = a = 0)$ and the alternative $(\delta \neq 0)$ when $a \in \mathscr{N}(C_\pi)$: the power of these tests cannot exceed the nominal levels. So, the practice of pretesting based on exogeneity tests is unreliable in this case.

Theorem **6.4** characterizes the distributions of the statistics in the special case of Gaussian errors.

Theorem 6.4 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Let the assumptions of Theorem* **6.2** *hold. If furthermore the normality assumption* (6.27) *holds and* $Z = [Z_1, Z_2]$ *is fixed, then*

$$
\begin{aligned}
\mathscr{H}_1 &= T[\mu_1 + \bar{v}]' \Gamma_1(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}], \\
\mathscr{H}_2 &= T[\mu_1 + \bar{v}]' \Gamma_2(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \sim T\phi_1(\bar{v}, \nu_1)/\phi_2(\bar{v}, \nu_3), \\
\mathscr{H}_3 | \bar{V} &\sim T/[1 + \kappa_2^{-1} F(T - k_1 - 2G, G; \upsilon_2, \nu_1)] \leq \bar{\kappa}_1^* F(G, T - k_1 - 2G; \nu_1, \upsilon_2), \\
\mathscr{T}_1 | \bar{V} &\sim F(G, k_2 - G; \nu_1, \upsilon_1), \ \mathscr{T}_2 | \bar{V} \sim F(G, T - k_1 - 2G; \nu_1, \upsilon_2), \\
\mathscr{T}_3 &= \kappa_2[\mu_1 + \bar{v}]' \Gamma_2(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \sim \kappa_2 \phi_1(\bar{v}, \nu_1)/\phi_2(\bar{v}, \nu_3), \\
\mathscr{T}_4 | \bar{V} &\sim \kappa_4/[1 + \kappa_2^{-1} F(T - k_1 - 2G, G; \upsilon_2, \nu_1)] \leq \bar{\kappa}_2^* F(G, T - k_1 - 2G; \nu_1, \upsilon_2), \\
\mathscr{R}\mathscr{H} | \bar{V} &\sim F(k_2, T - k - G; \nu_R, \upsilon_R),
\end{aligned}
$$

where $\phi_1(\bar{v}, \nu_1) | \bar{V} = [\mu_1 + \bar{v}]' C' \Delta^{-1} C [\mu_1 + \bar{v}] | \bar{V} \sim \chi^2(G; \nu_1)$, $\phi_2(\bar{v}, \nu_3) | \bar{V} = \omega_{IV}^2 | \bar{V} \sim \chi^2(T - k_1 - G; \nu_3)$, $\nu_1 = \mu_1' C' \Delta^{-1} C \mu_1$, $\nu_3 = \mu_1'(D_*' D^*)\mu_1$, $\upsilon_1 = \mu_1' E \mu_1$, $\upsilon_2 = \mu_1'(C_* - C'\Delta^{-1}C)\mu_1$, $\nu_R = \mu_1' P_{D_1 Z_2} \mu_1$, $\upsilon_R = \mu_1'(D_1 - P_{D_1 Z_2})\mu_1$, $\bar{\kappa}_1^* = TG/(T - k_1 - 2G)$, $\bar{\kappa}_2^* = (T - k_1 - G)G/(T - k_1 - 2G)$.

The above theorem entails that given $\bar{V}$, the statistics $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{R}\mathscr{H}$ follow double noncentral F-distributions, while $\mathscr{T}_4$ and $\mathscr{H}_3$ are bounded by a double noncentral F-type distribution. However, the distributions of $\mathscr{T}_3$, $\mathscr{H}_2$ and $\mathscr{H}_1$ cannot be characterized by standard distributions. As in Theorem **6.2**, μ_1 is the factor that determines power. If $\mu_1 \neq 0$, the exogeneity tests have power. However, when $\mu_1 = 0$, all tests have no power as shown in Corollary **6.5**.

Corollary 6.5 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem* **6.4**, *if* $a \in \mathscr{N}(C_\pi)$, *we have* $\nu_1 = \nu_3 = \upsilon_1 = \upsilon_2 = \nu_R = \upsilon_R = 0$ *so that*

$$
\begin{aligned}
\mathscr{H}_1 &= T\bar{v}' \Gamma_1(\mu_2, \bar{v}, \bar{V})\bar{v}, \ \mathscr{H}_2 = T\bar{v}' \Gamma_2(\mu_2, \bar{v}, \bar{V})\bar{v} \sim T\phi_1(\bar{v})/\phi_2(\bar{v}), \\
\mathscr{H}_3 &\sim T/(1 + \kappa_2^{-1} F(T - k_1 - 2G, G)) \leq \bar{\kappa}_1^* F(G, T - k_1 - 2G), \\
\mathscr{T}_1 &\sim F(G, k_2 - G), \ \mathscr{T}_2 \sim F(G, T - k_1 - 2G), \\
\mathscr{T}_3 &= \kappa_2 \bar{v}' \Gamma_2(\mu_2, \bar{v}, \bar{V})\bar{v} \sim \kappa_2 \phi_1(\bar{v})/\phi_2(\bar{v}), \\
\mathscr{T}_4 &\sim \kappa_4/[1 + \kappa_2^{-1} F(T - k_1 - 2G, G)] \leq \bar{\kappa}_2^* F(G, T - k_1 - 2G), \\
\mathscr{R}\mathscr{H} &\sim F(k_2, T - k - G),
\end{aligned}
$$

where $\phi_1(\bar{v}) \equiv \phi_1(\bar{v}, 0)$, $\phi_2(\bar{v}) \equiv \phi_2(\bar{v}, 0)$, $\phi_1(\bar{v}, \nu_1)$, $\phi_1(\bar{v}, \nu_3)$, $\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})$, $i = 1, 2$ *are defined in Theorem* **6.4**.

Observe that when $a \in \mathscr{N}(C_\pi)$, the non-centrality parameters in the F-distributions vanish. In particular, under the null hypothesis H_0, we have $a = 0 \in \mathscr{N}(C_\pi)$ and all exogeneity tests are pivotal. Furthermore, all exogeneity test statistics have the same distribution under the null hypothesis ($\delta = a = 0$) and the alternative ($\delta \neq 0$): the power of the tests cannot exceed the nominal levels.

We now describe the exact procedure for testing exogeneity even with non-Gaussian errors: the Monte Carlo exogeneity tests.

7. Exact Monte Carlo exogeneity (MCE) tests

The finite-sample characterization of the distribution of exogeneity test statistics in the previous section show that the tests are typically robust to weak instruments (level is controlled). However, these distributions (under the null hypothesis) of the statistics are not standard if the errors are non Gaussian. Furthermore, even for Gaussian errors, $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{T}_3$ cannot be characterized by standard distributions. This section develops exact Monte Carlo tests which are identification-robust even if the errors are non-Gaussian.

Consider again eq.(2.1) and assume that we test the strict exogeneity of Y, i.e. the hypothesis:

$$\mathrm{H}_0 \quad : \quad u \text{ is independent of } [Y, Z]\,. \tag{7.1}$$

If the distribution under H_0 of u/σ_u is given, the conditional distributions of exogeneity test statistics given $[Y, Z]$ are pivotal and therefore can be simulated [see Theorem **5.1**]. Let

$$\mathscr{W} \quad \in \quad \{\mathscr{H}_i, \mathscr{H}_l, \mathscr{R}\mathscr{H}, i = 1, 2, 3; l = 1, 2, 3, 4\}\,. \tag{7.2}$$

We shall consider two cases: first, the support of W is continuous and second, the support may be a discrete set.

We first focus on the case where exogeneity tests have continuous distributions. Let $W_1, \ldots, W_N$ be a sample of N replications of identically distributed exchangeable random variables with the same distribution as W [for more details on exchangeability, see Dufour (2006)]. Define $W(N) = (W_1, \ldots, W_N)'$ and let W_0 be the value of W based on the observed data. Let

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N + 1}, \tag{7.3}$$

$$\hat{G}_N(x) = \hat{G}_N[x; \mathscr{W}(N)], \tag{7.4}$$

where the survival function $\hat{G}_N$ is given by

$$\hat{G}_N[x; \mathscr{W}(N)] \quad = \quad \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[\mathscr{W}_i \geq x]}, \tag{7.5}$$

$$\begin{aligned} \mathbb{1}_C &= 1 \quad \text{if condition C holds,} \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{7.6}$$

Then, we can show that

$$P[\hat{p}_N(\mathscr{W}_0) \leq \alpha] \;=\; \frac{I[\alpha(N+1)]}{N+1} \quad \text{for } \; 0 \leq \alpha \leq 1\,, \tag{7.7}$$

[see Dufour (2006, Proposition 2.2)], where $I[x]$ is the largest integer less than or equal to x. So, $\hat{p}_N(W_0) \leq \alpha$ is the critical region of the MC-test with level $1-\alpha$ and $\hat{p}_N(W_0)$ is the MC-test p-value.

We shall now extend this procedure to the general case where the distribution of the statistic W may be discrete. Assume that $W(N) = (W_1, \ldots, W_N)'$ is a sequence of exchangeable random variables which may exhibit ties with positive probability. More precisely

$$P(\mathscr{W}_j = \mathscr{W}_{j'}) > 0 \quad \text{for } j \neq j',\, j, j' = 1, \ldots, N\,. \tag{7.8}$$

Let us associate each variable $W_j,\ j = 1, \ldots, N,$ with a random variable $U_j,\ j = 1, \ldots, N$ such that

$$\mathscr{U}_j, \ldots, \mathscr{U}_N \overset{i.i.d}{\sim} \mathscr{U}(0,1)\,, \tag{7.9}$$

$U(N) = (U_1, \ldots, U_N)'$ is independent of $W(N) = (W_1, \ldots, W_N)'$ where $U(0,1)$ is the uniform distribution on the interval $(0, 1)$. Then, we consider the pairs

$$\mathscr{Z}_j = (\mathscr{W}_j, \mathscr{U}_j),\ j = 1, \ldots, N\,, \tag{7.10}$$

which are ordered according to the lexicographic order:

$$(\mathscr{W}_j, \mathscr{U}_j) \leq (\mathscr{W}_{j'}, \mathscr{U}_{j'}) \Longleftrightarrow \{\mathscr{W}_j < \mathscr{W}_{j'} \; \text{ or } (\mathscr{W}_j = \mathscr{W}_{j'} \; \text{ and } \mathscr{U}_j \leq \mathscr{U}_{j'})\}\,. \tag{7.11}$$

Let us define the randomized p-value function as

$$\tilde{p}_N(x) = \frac{N\tilde{G}_N(x) + 1}{N+1}, \tag{7.12}$$

where the tail-area function $\tilde{G}_N$ is given by

$$\tilde{G}_N(x) = \tilde{G}_N[x; \mathscr{U}_0, \mathscr{W}(N), \mathscr{U}(N)], \tag{7.13}$$

and

$$\tilde{G}_N[x; \mathscr{U}_0, \mathscr{W}(N), \mathscr{U}(N)] \;=\; \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{[\mathscr{Z}_j \geq (x, \mathscr{U}_0)]}, \tag{7.14}$$

U_0 is a $U(0,1)$ random variable independent of $W(N)$ and $U(N)$. Then, we have

$$P[\tilde{p}_N(\mathscr{W}_0) \le \alpha] \;=\; \frac{I[\alpha(N+1)]}{N+1} \quad \text{for } 0 \le \alpha \le 1\,, \tag{7.15}$$

[see Dufour (2006, Proposition 2.4)]. So, $\tilde{p}_N(W_0) \le \alpha$ is the critical region of the MC-test with level $1-\alpha$ and $\tilde{p}_N(W_0)$ is the MC-test p-value.

The algorithm[3] for computing the Monte Carlo exogeneity tests p-values in the continuous distributions setup, is described as follows:

1. compute the test statistic W_0 based on the observed data;
2. generate $i.i.d.$ variables $u^{(j)} = [u_1^{(j)}, \ldots, u_T^{(j)}]'$, $\quad j = 1, \ldots, N$, according to the selected distribution–for example, $u_t^{(j)} \sim N[0,1]$ for all $t = 1, \ldots, T$ and $j = 1, \ldots, N$. Since the distribution of W under H_0 does not involve either β or γ, compute the pseudo-samples as functions of the OLS estimators $\hat{\beta}$ and $\hat{\gamma}$ from the observed data, i.e.

$$y_t^{(j)} = Y_t^{(j)\prime}\hat{\beta} + Z_{1t}^{(j)\prime}\hat{\gamma} + u_t^{(j)}, t = 1, \ldots, T, \, j = 1, \ldots, N\,, \tag{7.16}$$

   given the observed data Y and Z_1;
3. compute the corresponding test statistics $W^{(j)}$, $\; j = 1, \ldots, N$;
4. compute the MC p-value

$$\hat{p}_{MC} = \hat{p}_N[\mathscr{W}_0]; \tag{7.17}$$

5. reject the null hypothesis H_0 at level α_1 if $\hat{p}_{MC} \le \alpha_1$.

The following section present the Monte Carlo experiment.

8. Simulation experiment

In this section, we analyze the finite-sample behaviour (size and power) of DWH and RH tests through a Monte Carlo experiment allowing for the presence of non Gaussian errors (Cauchy-type errors). Two versions of the tests are considered: (i) the standard DWH and RH tests [see Wu (1973), Hausman (1978) and Revankar and Hartley (1973)]; and (ii) the exact Monte Carlo version of these tests, namely MCE-tests.

Now, we consider the model described by the following data generating process:

$$y = Y_1\beta_1 + Y_2\beta_2 + u, \quad (Y_1, Y_2) = (Z_2\Pi_{21}, Z_2\Pi_{22}) + (V_1, V_2), \tag{8.1}$$

[3] The algorithm can easily be generalized to discrete cases.

where Z_2 is a $T \times k_2$ matrix of instruments such that Z_{2t} follow $i.i.d$ $\mathrm{N}(0, I_{k_2})$ for $t = 1, \ldots, T$, Π_{21} and Π_{22} are vectors of dimension k_2. We assume that

$$u = Va + \varepsilon = V_1 a_1 + V_2 a_2 + \varepsilon, \tag{8.2}$$

where a_1 and a_2 are 2×1 vectors and ε is independent with $V = (V_1, V_2)$, V_1 and V_2 are $T \times 1$ vectors. Throughout this subsection, we consider two setups: (1) V_t and ε_t are independent such that

$$(V_{1t}, V_{2t})' \overset{i.i.d}{\sim} N\left(0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad \text{and} \quad \varepsilon_t \overset{i.i.d}{\sim} N(0, 1), \quad \text{for all} \quad t = 1, \ldots, T, \tag{8.3}$$

(2) V_t and ε_t are independent such that

$$V_{1t}, V_{2t} \quad \sim \quad \text{i.i.d standard Cauchy distribution} \tag{8.4}$$

$$\text{and} \quad \varepsilon_t \quad \sim \quad \text{i.i.d standard Cauchy distribution}, t = 1, \ldots, T. \tag{8.5}$$

We define

$$\Pi_{21} = \eta_1 C_0, \; \Pi_{22} = \eta_2 C_1, \tag{8.6}$$

where η_1 and η_2 take the value 0 (design of complete non identification), .01 (design of weak identification) or .5 (design of strong identification), $[C_0, C_1]$ is a $k_2 \times 2$ matrix obtained by taking the first two columns of the identity matrix of order k_2. Equation (8.6) allows us to consider partial identification of $\beta = (\beta_1, \beta_2)'$. In particular, if $\Pi_{21} = 0$ and $\Pi_{22}'\Pi_{22} \neq 0$, β_1 is not identified but β_2 is. The number of instruments k_2 varies in $\{5, 10, 20\}$ and the true value of β is set at $\beta_0 = (2, 5)'$. Note that when η_1 and η_2 belong to $\{0, .01\}$, the instruments Z_2 are weak and both ordinary least squares and two-stage least squares estimators of β in (8.1) are biased and inconsistent unless $a_1 = a_2 = 0$. The sample size is fixed at $T = 50$. The endogeneity parameter a is chosen such that

$$a = (a_1, a_2)' \in \{(-20, 0)', (-5, 5)', (0, 0)', (.5, .2)', (100, 100)'\}. \tag{8.7}$$

From the above notations, the usual exogeneity hypothesis of Y is expressed as

$$\mathrm{H}_0 : a = (a_1, a_2)' = (0, 0)'. \tag{8.8}$$

The nominal level of the tests for the standard DWH and RH tests is 5%. For each value of the parameter a, we compute the empirical rejection probability of all test statistics. When $a = 0$, the rejection frequencies are the empirical levels of the tests. However, when $a \neq 0$, the rejection frequencies represent the power of the tests. Section 8.1 presents the results for the standard DWH and RH tests.

8.1. Standard exogeneity tests

For the DWH and RH standard tests, the number of replications is set at $N = 10,000$. Table 1 presents the results when the errors are Gaussian while Table 2 contains those for the Cauchy-type errors. In the first column of each table, we report the statistics, while in the second column we report the values of k_2 (number of excluded instruments). In the other columns, for each value of the endogeneity parameter a and the quality of the instruments η_1 and η_2, the rejection frequencies at nominal level 5% are reported.

From the results of the tables, we then observe that:

1. all DWH and RH tests are identification-robust (level is controlled) whether the errors are Gaussian or not;

2. for Gaussian errors [setup (8.3)], $\mathscr{T}_1$, $\mathscr{T}_2$, $\mathscr{T}_4$, $\mathscr{H}_3$, and RH control the level while $\mathscr{T}_3$, $\mathscr{H}_1$ and $\mathscr{H}_2$ are conservative;

3. for Cauchy-type errors [setup (8.5)], unlike the previous case, in addition to $\mathscr{T}_3$, $\mathscr{H}_1$ and $\mathscr{H}_2$, $\mathscr{T}_1$ is also conservative when identification is deficient. The results are the same as in setup (8.3) for $\mathscr{T}_2$, $\mathscr{T}_4$, $\mathscr{H}_3$, and RH;

4. all exogeneity tests exhibit power even if not all parameters are identified, provided partial identification holds. Hence, the results of Staiger and Stock (1997) and Guggenberger (2010) may be misleading;

5. when identification is completely deficient, *i.e.* $\eta_1 = \eta_2 = 0$ (irrelevant instruments), all DWH and RH tests have no power whether the errors are Gaussian or not [similar to Staiger and Stock (1997) and Guggenberger (2010)];

6. in terms of power comparison, $\mathscr{H}_3$ dominates $\mathscr{H}_2$ and $\mathscr{H}_2$ dominates $\mathscr{H}_1$ irrespective of whether identification is deficient or not. In the same way, $\mathscr{T}_2$ dominates T_4, $\mathscr{T}_4$ dominates $\mathscr{T}_1$ and $\mathscr{T}_1$ dominates T_3.

Table 1. Power of exogeneity tests at nominal level 5%; $G = 2$, $T = 50$

	k_2	$(a_1,a_2)' = (-20,0)'$			$(a_1,a_2)' = (-5,5)'$			$(a_1,a_2)' = (0,0)'$			$(a_1,a_2)' = (.5,.2)'$			$(a_1,a_2)' = (100,100)'$		
		$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$
$\mathscr{T}_1$	5	4.98	4.6	65.81	5.26	4.92	70.9	4.87	5.06	5.24	5.09	4.84	19.85	4.94	4.18	70.09
$\mathscr{T}_2$	5	4.98	24.92	100	5.04	6.77	100	4.96	5.38	5.26	4.87	4.61	53.19	4.91	76.71	100
$\mathscr{T}_3$	5	0	0.19	97.93	0.02	0.05	97.85	0.02	0.03	0.59	0.03	0	29.02	0.01	5.83	97.93
$\mathscr{T}_4$	5	4.64	24.07	100	4.67	6.29	100	4.63	4.91	4.93	4.51	4.42	52	4.62	76.25	100
$\mathscr{H}_1$	5	0	0.09	92.53	0.01	0.02	91.83	0.01	0.02	0.26	0	0	17.97	0	3.59	92.48
$\mathscr{H}_2$	5	0.01	0.25	98.09	0.03	0.05	98.02	0.02	0.04	0.74	0.04	0	31.42	0.02	6.89	98.14
$\mathscr{H}_3$	5	5.34	25.73	100	5.33	7.19	100	5.27	5.72	5.56	5.18	4.92	54.41	5.31	77.11	100
$\mathscr{RH}$	5	4.84	45.25	100	5.36	7.83	100	5.04	5.2	4.9	4.88	4.73	41.31	5.02	100	100
$\mathscr{T}_1$	10	4.9	3.95	98.38	4.92	5.34	98.93	4.82	4.81	5.25	4.88	5.22	34.18	4.91	3.28	99.23
$\mathscr{T}_2$	10	5.01	17.5	100	5.19	6.2	100	5.16	4.88	5.07	4.77	5.45	54.24	4.8	50.74	100
$\mathscr{T}_3$	10	0.35	1.88	100	0.38	0.29	100	0.3	0.33	1.47	0.36	0.3	43.01	0.22	14.7	100
$\mathscr{T}_4$	10	4.65	16.77	100	4.75	5.73	100	4.78	4.55	4.72	4.45	5.02	52.81	4.46	50.05	100
$\mathscr{H}_1$	10	0.16	1.05	99.31	0.18	0.14	99.22	0.2	0.14	0.49	0.14	0.14	28.92	0.1	9.88	99.25
$\mathscr{H}_2$	10	0.46	2.3	100	0.48	0.42	100	0.38	0.43	1.76	0.46	0.39	45.54	0.33	16.85	100
$\mathscr{H}_3$	10	5.32	18.11	100	5.43	6.56	100	5.46	5.18	5.41	5.06	5.75	55.31	5.12	51.25	100
$\mathscr{RH}$	10	5.17	57.58	100	4.83	7.62	100	4.83	5.34	4.97	4.93	5.41	34.5	4.57	100	100
$\mathscr{T}_1$	20	4.93	2.26	99.8	4.94	4.64	99.78	4.9	5.02	5.07	5.02	4.93	39.4	5.02	1.5	99.96
$\mathscr{T}_2$	20	4.75	8.97	100	4.9	5.54	100	5.09	5.32	4.99	4.95	4.94	49.34	4.92	17.32	100
$\mathscr{T}_3$	20	1.95	3.73	100	1.82	2.01	100	2.1	2.02	2.79	2.01	1.95	44.9	1.94	9.2	100
$\mathscr{T}_4$	20	4.43	8.42	100	4.51	5.21	100	4.74	5.04	4.61	4.63	4.57	47.89	4.52	16.45	100
$\mathscr{H}_1$	20	1.08	2.43	99.89	1.13	1.08	99.82	1.13	1.2	1.03	1.08	1.21	29.88	1.15	6.44	99.7
$\mathscr{H}_2$	20	2.32	4.37	100	2.26	2.6	100	2.67	2.57	3.28	2.46	2.48	47.46	2.33	10.39	100
$\mathscr{H}_3$	20	5.15	9.36	100	5.25	5.73	100	5.4	5.68	5.41	5.23	5.18	50.31	5.23	17.76	100
$\mathscr{RH}$	20	4.88	79.08	100	5.03	8.36	100	5.38	5	5.21	5.07	5.04	24.88	5.3	100	100

Table 1 (continued). Power of exogeneity tests at nominal level 5%; $G = 2$, $T = 50$

	k_2	$(a_1, a_2)' = (-20, 0)'$			$(a_1, a_2)' = (-5, 5)'$			$(a_1, a_2)' = (0, 0)'$			$(a_1, a_2)' = (.5, .2)'$			$(a_1, a_2)' = (100, 100)'$		
		$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$
$\mathscr{T}_1$	5	4.73	15.16	81.58	69.69	68.76	78.22	4.91	5.26	5	8.01	7.48	24.2	63.6	65.14	78.04
$\mathscr{T}_2$	5	5.1	37.9	100	100	100	100	5.51	5.29	5.2	12.95	12.42	64.31	100	100	100
$\mathscr{T}_3$	5	0.63	18.25	98.68	98.15	98.26	98.50	0.75	0.85	0.83	3.82	3.47	42.79	97.43	97.09	98.52
$\mathscr{T}_4$	5	4.77	36.89	100	100	100	100	5.06	4.98	4.78	12.24	11.72	63.06	100	100	100
$\mathscr{H}_1$	5	0.27	10.48	90.44	92	92.3	92.20	0.39	0.29	0.32	1.93	1.69	24.39	92.4	91.95	92.12
$\mathscr{H}_2$	5	0.77	20.16	98.82	98.33	98.43	98.52	0.87	0.96	0.99	4.44	4.08	45.64	97.59	97.31	98.64
$\mathscr{H}_3$	5	5.48	38.88	100	100	100	100	5.83	5.64	5.41	13.39	12.95	65.44	100	100	100
$\mathscr{RH}$	5	5.13	28.27	100	100	100	100	4.77	5.13	5.17	9.81	10.28	50.59	100	100	100
$\mathscr{T}_1$	10	5.18	26.81	99.76	98.81	99.17	99.56	5.26	5.3	4.86	11.05	11.61	43.71	99.12	99.28	99.74
$\mathscr{T}_2$	10	5.29	41.58	100	100	100	100	4.92	5.19	5.07	13.49	14.75	66.24	100	100	100
$\mathscr{T}_3$	10	1.7	31.1	99.98	99.97	99.99	100	1.58	1.6	1.88	7.75	8.29	57.52	100	100	100
$\mathscr{T}_4$	10	4.96	40.35	100	100	100	100	4.57	4.87	4.67	12.81	14	65.15	100	100	100
$\mathscr{H}_1$	10	0.73	18.21	98.22	99.08	98.98	98.9	0.55	0.5	0.48	3.34	3.88	32.85	99.28	99.26	98.29
$\mathscr{H}_2$	10	2	33.67	99.98	99.98	100	100	1.88	2.03	2.31	8.65	9.3	60.4	100	100	100
$\mathscr{H}_3$	10	5.61	42.64	100	100	100	100	5.3	5.53	5.38	14.05	15.32	67.3	100	100	100
$\mathscr{RH}$	10	5.24	24.16	100	100	100	100	4.92	5.07	5.11	8.55	8.94	43.87	100	100	100
$\mathscr{T}_1$	20	5.12	27.67	99.96	99.45	99.48	99.62	4.86	4.91	4.29	10.45	10.95	41.15	99.91	99.9	99.94
$\mathscr{T}_2$	20	5.06	34.7	100	100	100	100	4.93	4.77	4.3	11.85	12.03	51.76	100	100	100
$\mathscr{T}_3$	20	2.97	30.26	100	100	100	100	3.2	2.88	2.74	9.14	9.14	47.52	100	100	100
$\mathscr{T}_4$	20	4.7	33.32	100	100	100	100	4.57	4.45	3.97	11.13	11.34	50.35	100	100	100
$\mathscr{H}_1$	20	1.2	17.73	99.24	99.93	99.91	99.93	1.1	1.03	0.72	4.51	4.53	27.81	99.77	99.81	98.75
$\mathscr{H}_2$	20	3.59	32.57	100	100	100	100	3.65	3.39	3.27	10.24	10.25	50.07	100	100	100
$\mathscr{H}_3$	20	5.32	35.69	100	100	100	100	5.25	5.06	4.55	12.42	12.55	52.91	100	100	100
$\mathscr{RH}$	20	5.46	16.17	100	100	100	100	5.2	4.64	4.82	7.45	7.45	26.62	100	100	100

Table 2. Power of exogeneity tests at nominal level 5% with Cauchy errors; $G=2$, $T=50$

	k_2	$(a_1,a_2)'=(-20,0)'$			$(a_1,a_2)'=(-5,5)'$			$(a_1,a_2)'=(0,0)'$			$(a_1,a_2)'=(.5,.2)'$			$(a_1,a_2)'=(100,100)'$		
		$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$
$\mathscr{T}_1$	5	4.96	4.94	4.9	4.98	4.95	5.14	5.19	5.23	4.97	5.1	5.34	5.23	5.01	4.83	6.66
$\mathscr{T}_2$	5	5.08	8.58	59.38	5.48	6.02	24.51	5.46	5.38	5.29	5.15	4.97	5.54	5.32	44.68	81.16
$\mathscr{T}_3$	5	0.05	0.08	4.91	0.02	0.03	0.65	0	0.03	0.05	0.02	0.01	0.02	0.01	1.62	8.71
$\mathscr{T}_4$	5	4.82	8.08	58.8	5.19	5.62	23.87	5.14	5	4.9	4.75	4.63	5.14	5.01	44	80.76
$\mathscr{H}_1$	5	0.04	0.02	3.26	0.01	0	0.33	0	0.01	0.02	0.01	0	0.01	0	0.91	6.2
$\mathscr{H}_2$	5	0.07	0.11	5.86	0.05	0.04	0.81	0	0.05	0.05	0.02	0.03	0.03	0.02	2.02	9.95
$\mathscr{H}_3$	5	5.41	9.01	59.84	5.81	6.38	25.21	5.67	5.7	5.64	5.49	5.23	5.77	5.57	45.34	81.48
$\mathscr{RH}$	5	5.13	12.29	82.91	5.61	6.79	40.66	6.04	5.98	5.93	4.88	4.43	5.06	6.12	68.34	96.73
$\mathscr{T}_1$	10	5.61	4.79	5.07	4.97	5.2	4.63	4.83	5.48	4.7	5.04	5.08	5.22	5.09	2.95	3.24
$\mathscr{T}_2$	10	5.42	6.48	38.72	5.53	5.41	9.28	4.79	5.17	4.81	4.92	4.94	5.14	5.01	22.57	54.36
$\mathscr{T}_3$	10	0.39	0.44	10.96	0.3	0.3	0.8	0.31	0.28	0.32	0.34	0.28	0.19	0.38	3.53	18.51
$\mathscr{T}_4$	10	5.08	6.09	38.06	5.24	5	8.86	4.45	4.87	4.46	4.61	4.63	4.83	4.69	21.74	53.51
$\mathscr{H}_1$	10	0.17	0.17	7.6	0.11	0.13	0.42	0.16	0.08	0.09	0.14	0.17	0.09	0.14	2.06	13.04
$\mathscr{H}_2$	10	0.49	0.65	12.56	0.46	0.42	1.11	0.4	0.38	0.45	0.44	0.39	0.33	0.51	4.19	20.96
$\mathscr{H}_3$	10	5.61	6.8	39.32	5.8	5.64	9.66	5.01	5.42	5.05	5.19	5.2	5.43	5.33	23.16	55.1
$\mathscr{RH}$	10	6.09	11.71	81.63	6.41	5.77	22.73	5.06	4.67	4.98	4.22	4.63	4.77	3.86	62.53	96.32
$\mathscr{T}_1$	20	5.27	5.02	3.63	4.64	4.63	4.35	4.96	5.27	5.09	4.77	5.16	4.85	5.1	3	2.55
$\mathscr{T}_2$	20	5.34	5.4	13.09	4.94	4.9	6.76	4.85	5.06	4.98	4.76	5.26	4.98	4.84	8.73	18.56
$\mathscr{T}_3$	20	2.03	2.16	6.97	1.8	1.77	2.45	1.95	2.09	1.87	1.91	2.19	1.88	2.06	3.74	11.08
$\mathscr{T}_4$	20	5.03	5.13	12.58	4.6	4.57	6.42	4.47	4.68	4.67	4.48	5.01	4.7	4.57	8.2	18.04
$\mathscr{H}_1$	20	1.21	1.25	4.78	1.05	1.01	1.61	1.14	1.19	1.06	0.94	1.35	1.09	1.26	2.42	8.21
$\mathscr{H}_2$	20	2.54	2.62	8.03	2.27	2.25	3.25	2.3	2.62	2.33	2.35	2.56	2.4	2.43	4.38	12.28
$\mathscr{H}_3$	20	5.72	5.69	13.49	5.14	5.12	7.07	5.21	5.4	5.29	5.04	5.5	5.27	5.08	9.06	19.12
$\mathscr{RH}$	20	6.3	9.15	75.83	4.05	4.15	23.42	6.55	6.42	6.83	5.49	5.03	5.27	5.01	54.94	94.83

Table 2 (continued). Power of exogeneity tests at nominal level 5% with Cauchy errors; $G = 2$, $T = 50$

	k_2	$(a_1,a_2)' = (-20,0)'$			$(a_1,a_2)' = (-5,5)'$			$(a_1,a_2)' = (0,0)'$			$(a_1,a_2)' = (.5,.2)'$			$(a_1,a_2)' = (100,100)'$		
		$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$
$\mathscr{T}_1$	5	5.21	4.86	4.43	4.53	5.03	4.65	4.88	4.42	4.88	5.31	5.16	5.05	4.64	4.66	5.31
$\mathscr{T}_2$	5	4.51	7.01	52.88	22.05	22.32	33.84	4.83	4.77	5.01	5.02	4.91	5.52	77.26	77.11	78.14
$\mathscr{T}_3$	5	0.01	0.01	3.51	0.72	0.6	1.47	0.01	0.01	0.04	0.01	0.06	0.02	8.81	9.01	10.27
$\mathscr{T}_4$	5	4.24	6.6	52.29	21.41	21.56	33.16	4.42	4.5	4.7	4.64	4.63	5.07	76.83	76.71	77.77
$\mathscr{H}_1$	5	0	0	2.22	0.42	0.21	0.95	0	0	0.01	0	0.02	0.01	6.38	6.37	7.38
$\mathscr{H}_2$	5	0.04	0.01	4.22	0.89	0.75	1.74	0.02	0.03	0.05	0.02	0.06	0.03	10.02	10.14	11.51
$\mathscr{H}_3$	5	4.84	7.31	53.53	22.61	22.8	34.49	5.19	5.08	5.42	5.41	5.24	5.77	77.6	77.53	78.63
$\mathscr{RH}$	5	4.36	8.62	77.57	37.03	36.83	53.81	5.32	5.34	5.36	5.03	5.29	5.46	96.24	96.42	97.63
$\mathscr{T}_1$	10	4.72	4.97	4.34	4.87	5.41	5.3	5.2	5.3	5.16	4.89	4.93	4.7	5.07	4.59	4.81
$\mathscr{T}_2$	10	4.53	6.71	36.17	13.87	13.91	17.44	4.94	5.01	5.11	5.11	5.14	5.15	49.25	49.57	52.89
$\mathscr{T}_3$	10	0.23	0.49	10.23	1.6	1.95	3.09	0.34	0.34	0.27	0.27	0.34	0.31	16.39	15.82	18.7
$\mathscr{T}_4$	10	4.16	6.3	35.3	13.24	13.31	16.82	4.65	4.68	4.7	4.77	4.73	4.85	48.54	48.81	52.05
$\mathscr{H}_1$	10	0.08	0.25	7.01	0.9	1.04	1.86	0.12	0.19	0.15	0.08	0.09	0.11	12.12	11.64	13.86
$\mathscr{H}_2$	10	0.34	0.75	11.8	2.03	2.38	3.62	0.44	0.43	0.35	0.42	0.49	0.41	18.12	17.91	20.72
$\mathscr{H}_3$	10	4.91	7.18	36.81	14.51	14.37	18.09	5.17	5.45	5.45	5.37	5.44	5.46	49.86	50.25	53.49
$\mathscr{RH}$	10	4.94	9.41	78.79	34.19	33.03	45.3	5.36	4.98	5.44	5.11	5.01	5.46	95.77	95.26	97.22
$\mathscr{T}_1$	20	4.83	4.39	2.6	4.31	4.21	3.47	4.85	5.12	4.67	4.66	4.85	5.05	2.26	2.19	1.79
$\mathscr{T}_2$	20	4.61	4.6	13.11	6.41	6.08	6.78	4.65	4.85	4.95	4.56	4.7	5.13	18.38	17.85	18.44
$\mathscr{T}_3$	20	2.04	1.85	6.7	2.6	2.54	3	1.69	1.99	1.9	1.88	2	2.23	11.17	10.59	10.62
$\mathscr{T}_4$	20	4.21	4.34	12.41	6.09	5.79	6.48	4.27	4.57	4.73	4.23	4.4	4.8	17.78	17.22	17.83
$\mathscr{H}_1$	20	1.12	1.16	4.61	1.59	1.55	1.66	1.01	1.07	1.12	1.08	1.15	1.35	8.44	7.93	7.45
$\mathscr{H}_2$	20	2.44	2.2	7.67	3.04	3.12	3.52	2.16	2.48	2.39	2.26	2.41	2.76	12.37	11.67	12.04
$\mathscr{H}_3$	20	4.86	4.93	13.56	6.75	6.36	7.23	4.93	5.17	5.26	4.85	4.97	5.5	19.04	18.46	18.95
$\mathscr{RH}$	20	6.64	9.64	75.85	18.22	18.08	33.69	5.31	5.11	5.31	4.38	4.64	4.93	94.4	94.26	96.09

We now focus on the exact Monte Carlo exogeneity tests.

8.2. Exact Monte Carlo exogeneity (MCE) tests

In this we present we study the properties of the exact tests following the algorithm described in Section 7. The DGP is the same as described above except for the distribution of the errors. We consider three types of errors: Gaussian, Cauchy and Student. Tables 3- 5 present the results for $M = 99$ replications.

We note that unlike the standard exogeneity tests, the level is controlled in all cases, as expected. Furthermore, the power of all tests has improved, in particular in Cauchy and Student distributions setups (Tables 4-5). As the standard versions of the tests, the MCE tests exhibits power provided the partial identification. But power do not exist when all instruments are weak. In addition, we observe that all MCE tests perform better (in terms of power) in Gaussian and Student distributions setups than the Cauchy distribution one.

Overall, our results clearly suggest that finite-sample improvement of standard exogeneity tests is feasible, whether the errors are Gaussian and the identification is strong or not. Hence, the conclusion by Staiger and Stock (1997) that size adjustment is infeasible may be misleading.

Table 3 . Power of MCE tests with Gaussian errors

	k_2	$(a_1,a_2)'=(-20,0)'$			$(a_1,a_2)'=(-5,5)'$			$(a_1,a_2)'=(0,0)'$			$(a_1,a_2)'=(.5,.2)'$			$(a_1,a_2)'=(100,100)'$		
		$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$
$\mathscr{T}_1$	5	5	7	30	4	4	41	4	5	5	12	9	13	77	78	81
$\mathscr{T}_2$	5	5	8	100	5	6	99	5	4	5	13	7	23	100	100	100
$\mathscr{T}_3$	5	3	9	93	3	4	95	4	3	4	12	8	22	91	94	89
$\mathscr{T}_4$	5	5	8	100	5	5	99	5	5	5	13	7	23	100	100	100
$\mathscr{H}_1$	5	4	9	91	3	4	94	4	4	4	12	9	22	91	93	89
$\mathscr{H}_2$	5	4	9	93	3	4	95	4	5	5	12	8	22	91	94	89
$\mathscr{H}_3$	5	5	8	100	5	6	99	5	5	5	13	7	23	100	100	100
$\mathscr{RH}$	5	5	18	100	5	6	100	5	5	5	13	9	23	100	100	100
$\mathscr{T}_1$	10	4	7	55	4	3	50	4	5	5	6	5	17	97	95	95
$\mathscr{T}_2$	10	5	8	99	5	7	99	5	5	5	6	8	23	99	100	98
$\mathscr{T}_3$	10	4	5	99	4	3	99	3	4	3	5	7	20	99	98	97
$\mathscr{T}_4$	10	5	8	99	5	7	99	5	5	5	6	8	23	99	100	98
$\mathscr{H}_1$	10	3	5	99	3	3	98	3	4	3	6	7	21	99	98	96
$\mathscr{H}_2$	10	4	5	99	4	3	99	4	3	4	5	7	22	99	98	97
$\mathscr{H}_3$	10	5	8	99	5	7	99	5	5	5	6	8	23	99	100	98
$\mathscr{RH}$	10	5	16	100	5	6	100	4	5	4	6	13	24	100	100	100
$\mathscr{T}_1$	20	5	6	33	4	5	68	4	5	4	6	7	10	88	83	90
$\mathscr{T}_2$	20	5	7	80	5	7	99	5	4	5	6	8	12	92	84	93
$\mathscr{T}_3$	20	4	7	82	3	4	99	3	3	3	5	8	10	94	88	93
$\mathscr{T}_4$	20	5	7	80	5	6	99	4	5	4	6	8	12	92	84	93
$\mathscr{H}_1$	20	3	6	81	4	4	99	3	3	3	6	7	10	94	89	93
$\mathscr{H}_2$	20	3	7	82	3	4	99	3	3	3	5	8	11	94	88	93
$\mathscr{H}_3$	20	4	7	80	5	6	99	5	5	5	6	8	12	92	84	93
$\mathscr{RH}$	20	5	6	100	4	7	100	3	5	4	12	13	12	100	100	100

Table 4 . Power of MCE tests with Cauchy errors

	k_2	$(a_1,a_2)'=(-20,0)'$			$(a_1,a_2)'=(-5,5)'$			$(a_1,a_2)'=(0,0)'$			$(a_1,a_2)'=(.5,.2)'$			$(a_1,a_2)'=(100,100)'$		
		$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$
$\mathscr{T}_1$	5	4	7	27	4	9	17	2	4	3	6	5	5	38	41	40
$\mathscr{T}_2$	5	5	8	62	5	6	32	5	5	4	9	6	5	76	81	78
$\mathscr{T}_3$	5	4	5	38	5	5	20	5	3	3	7	4	4	50	50	50
$\mathscr{T}_4$	5	5	8	62	5	6	32	5	5	4	9	6	5	76	81	78
$\mathscr{H}_1$	5	4	5	38	4	5	20	5	5	3	7	3	4	47	51	49
$\mathscr{H}_2$	5	4	5	38	4	5	20	5	3	3	7	4	4	50	50	50
$\mathscr{H}_3$	5	5	8	62	5	6	32	5	5	4	9	6	5	76	81	78
$\mathscr{RH}$	5	5	7	82	5	5	47	5	5	5	8	6	6	95	96	98
$\mathscr{T}_1$	10	3	2	24	5	8	10	3	5	4	4	6	5	34	36	37
$\mathscr{T}_2$	10	5	3	37	5	6	15	4	3	5	6	9	6	55	54	51
$\mathscr{T}_3$	10	5	5	34	4	3	12	4	2	4	4	7	5	43	49	44
$\mathscr{T}_4$	10	5	3	37	5	6	15	4	3	5	6	9	4	55	54	51
$\mathscr{H}_1$	10	4	4	34	4	3	14	4	2	4	4	7	4	43	52	41
$\mathscr{H}_2$	10	5	5	34	4	3	12	4	2	4	4	7	5	43	49	44
$\mathscr{H}_3$	10	5	3	37	5	6	15	4	3	5	6	9	6	55	54	51
$\mathscr{RH}$	10	5	7	82	4	5	37	4	4	5	7	10	7	96	96	93
$\mathscr{T}_1$	20	5	6	11	4	7	6	4	4	4	6	6	4	9	12	10
$\mathscr{T}_2$	20	5	9	15	5	7	8	5	5	4	6	5	5	14	19	12
$\mathscr{T}_3$	20	5	8	13	5	8	8	4	5	4	5	5	5	15	17	11
$\mathscr{T}_4$	20	5	9	15	5	7	8	5	5	4	6	5	5	14	19	12
$\mathscr{H}_1$	20	5	8	15	4	8	8	3	5	4	4	5	5	14	17	12
$\mathscr{H}_2$	20	5	8	13	5	8	8	4	5	4	5	5	5	15	17	11
$\mathscr{H}_3$	20	5	9	15	5	7	8	5	5	4	6	5	5	14	19	12
$\mathscr{RH}$	20	4	6	73	5	10	31	5	4	5	5	7	8	98	94	98

Table 5 . Power of MCE tests with Student errors

	k_2	$(a_1,a_2)'=(-20,0)'$			$(a_1,a_2)'=(-5,5)'$			$(a_1,a_2)'=(0,0)'$			$(a_1,a_2)'=(.5,.2)'$			$(a_1,a_2)'=(100,100)'$		
		$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=0$	$\eta_1=.01$ $\eta_2=0$	$\eta_1=.5$ $\eta_2=0$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$	$\eta_1=0$ $\eta_2=.5$	$\eta_1=.01$ $\eta_2=.5$	$\eta_1=.5$ $\eta_2=.5$
$\mathscr{T}_1$	5	3	6	24	6	7	33	5	4	4	7	7	9	75	68	63
$\mathscr{T}_2$	5	5	7	97	4	5	92	5	4	5	10	10	10	98	100	99
$\mathscr{T}_3$	5	4	4	86	6	6	72	4	3	3	10	8	6	89	82	90
$\mathscr{T}_4$	5	5	7	97	4	5	92	5	5	5	10	10	10	98	100	99
$\mathscr{H}_1$	5	3	4	86	6	6	71	3	3	4	10	8	6	87	82	90
$\mathscr{H}_2$	5	4	5	86	6	6	72	4	3	3	10	8	6	89	82	90
$\mathscr{H}_3$	5	5	5	97	4	5	92	5	4	4	10	10	10	98	100	99
$\mathscr{RH}$	5	5	7	100	4	8	100	5	5	4	12	12	10	100	100	100
$\mathscr{T}_1$	10	2	6	18	3	5	29	4	5	4	5	2	4	52	62	59
$\mathscr{T}_2$	10	4	7	90	5	5	82	5	5	5	6	4	8	90	90	93
$\mathscr{T}_3$	10	4	5	83	5	2	77	5	4	3	6	3	8	84	86	88
$\mathscr{T}_4$	10	4	7	90	5	5	82	5	3	5	6	4	8	90	90	93
$\mathscr{H}_1$	10	3	6	82	5	2	78	4	3	3	6	3	8	82	86	88
$\mathscr{H}_2$	10	4	6	83	5	2	77	3	3	4	6	3	8	84	86	88
$\mathscr{H}_3$	10	4	7	90	5	5	82	5	5	5	6	4	8	90	90	93
$\mathscr{RH}$	10	3	7	100	5	8	99	5	4	4	6	7	10	100	100	100
$\mathscr{T}_1$	20	4	6	15	4	10	15	5	4	5	5	4	4	50	51	55
$\mathscr{T}_2$	20	5	8	65	5	10	44	5	5	5	5	6	6	64	60	73
$\mathscr{T}_3$	20	3	6	65	4	11	48	4	4	3	5	7	4	64	61	74
$\mathscr{T}_4$	20	5	8	65	5	10	44	5	5	4	5	6	6	64	60	73
$\mathscr{H}_1$	20	4	5	65	4	11	48	3	3	2	5	7	4	64	61	74
$\mathscr{H}_2$	20	3	6	65	4	11	48	4	4	3	5	7	4	64	61	74
$\mathscr{H}_3$	20	5	8	65	5	10	44	5	5	5	5	6	6	64	60	73
$\mathscr{RH}$	20	5	7	100	5	14	98	4	4	5	5	9	5	100	100	100

9. Conclusion

This paper develops a finite-sample analysis of the distribution of the standard Durbin-Wu-Hausman and Revankar-Hartley specification tests under both the null hypothesis of exogeneity (level) and the alternative hypothesis of endogeneity (power), with or without identification. Our analysis provides several new insights and extensions of earlier procedures. The characterization of the finite-sample distributions of the statistics under the null hypothesis shows that all tests are typically robust to weak instruments (level is controlled). We provide a characterization of the power of the tests that clearly exhibits the factors that determine power. We show that exogeneity tests have no power in the extreme case where all IVs are weak [similar to Staiger and Stock (1997), and Guggenberger (2010)], but do have power as soon as we have one strong instrument. As a result, exogeneity tests can detect an exogeneity problem even if not all model parameters are identified, provided partial identification holds. Moreover, the finite-sample characterization of the distributions of the tests allows the construction of exact identification-robust exogeneity tests even in cases where conventional asymptotic theory breaks down. In particular, DWH and RH tests are valid even if the distribution of the errors does not have moments (Cauchy-type distribution, for example). We present a Monte Carlo experiment which confirms our finite-sample theory. The large-sample properties of the tests and estimation issues related to pretesting are examined in Doko and Dufour (2011).

A. Notes

A.1. Unified formulation of DWH test statistics

We establish the unified formulation of Durbin-Wu statistics in (4.1) - (3.10), as well as the three versions of Hausman (1978) statistic. From Wu (1973, Eqs. (2.1), (2.18), (3.16), (3.20)), T_l, $l = 1, 2, 3, 4$ are defined as

$$\begin{aligned}
\mathcal{T}_1 &= \kappa_1 Q^*/Q_1,\ \mathcal{T}_2 = \kappa_2 Q^*/Q_2,\ \mathcal{T}_1 = \kappa_3 Q^*/Q_3,\ \mathcal{T}_2 = \kappa_4 Q^*/Q_4, & \text{(A.1)}\\
Q^* &= (b_1 - b_2)' \left[(Y'A_2Y)^{-1} - (Y'A_1Y)^{-1}\right]^{-1}(b_1 - b_2), & \text{(A.2)}\\
Q_1 &= (y - Yb_2)'A_2(y - Yb_2),\ Q_2 = Q_4 - Q^*, & \text{(A.3)}\\
Q_4 &= (y - Yb_1)'A_1(y - Yb_1),\ Q_3 = (y - Yb_2)'A_1(y - Yb_2), & \text{(A.4)}\\
b_i &= (Y'A_iY)^{-1}Y'A_iy,\ i = 1, 2,\ A_1 = M_1,\ A_2 = M - M_1\,, & \text{(A.5)}
\end{aligned}$$

where b_1 is the ordinary least squares estimator of β, and b_2 is the instrumental variables method estimator of β. So, from our notations, $b_1 \equiv \hat{\beta}$ and $b_2 \equiv \tilde{\beta}$.

So, from (3.8) - (3.10), we have

$$Q^* = T\,(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\tilde{\sigma}^2(\tilde{\beta} - \hat{\beta})'\hat{\Sigma}_2^{-1}(\tilde{\beta} - \hat{\beta}), \quad \text{(A.6)}$$

$$Q_1 = T\,\tilde{\sigma}_1^2, \quad Q_3 = T\,\tilde{\sigma}^2, \quad Q_4 = T\,\hat{\sigma}^2\,, \quad \text{(A.7)}$$

$$Q_2 = Q_4 - Q^* = T\hat{\sigma}^2 - T\,(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\,\tilde{\sigma}_2^2 \quad \text{(A.8)}$$

so that $\mathcal{T}_l$, can be expressed as:

$$\mathcal{T}_l \;=\; \kappa_l(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta} - \hat{\beta})\,, \quad l = 1, 2, 3, 4\,, \quad \text{(A.9)}$$

where κ_l, and $\tilde{\Sigma}_l$ are defined in (4.1) - (3.10). The formulation in (A.9) shows clearly the link between Wu (1973) tests and Hausman (1978) test.

A.2. Regression interpretation of DWH test statistics

Consider Equations (**??**) - (3.5). First, we note that H_0 and H_b can be written as

$$\begin{aligned}
\mathrm{H}_0 &: R\theta = 0 \Leftrightarrow Rb = a\,,\\
H_b &: R_*\theta_* = 0 \Leftrightarrow R_*\theta_* = \beta - a\,,
\end{aligned}$$

where $R = \begin{bmatrix} 0 & 0 & I_G \end{bmatrix}$ and $R_* = \begin{bmatrix} I_G & 0 & -I_G \end{bmatrix}$. By definition, we have $\hat{\theta}_* = [\tilde{\beta}', \tilde{\gamma}', \tilde{b}']'$ and $\hat{\theta}_{*0} = [\hat{\beta}', \hat{\gamma}', \hat{\beta}']'$, where $\tilde{\beta}$ and $\tilde{\gamma}$ are the 2SLS estimators of β and γ and $\hat{\beta}$ and $\hat{\gamma}$ are the OLS estimators of β and γ based on the following model:

$$y = Y\beta + Z_1\gamma + u, \hat{Y} = Z\hat{\Pi},$$

with $\hat{\Pi} = (Z'Z)^{-1}Z'Y$. So, we can observe that

$$\begin{aligned} \hat{\theta}_{*0} &= \hat{\theta}_* + (\hat{X}'\hat{X})^{-1}R_*' \left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(-R_*\hat{\theta}_*) \\ S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) &= (\hat{\theta}_{*0} - \hat{\theta}_*)'\hat{X}'\hat{X}(\hat{\theta}_{*0} - \hat{\theta}_*) = (R_*\hat{\theta}_*)' \left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*). \end{aligned}$$

Furthermore, we have

$$\begin{aligned} R_*\hat{\theta} &= \begin{bmatrix} I_G & 0 & -I_G \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\gamma} \\ \tilde{b} \end{bmatrix} = \tilde{\beta} - \tilde{b}, \\ \hat{X}'\hat{X} &= \begin{bmatrix} (\hat{X}_1'\hat{X}_1) & 0 \\ 0 & (\hat{V}'\hat{V}) \end{bmatrix}, (\hat{X}'\hat{X})^{-1} = \begin{bmatrix} (\hat{X}_1'\hat{X}_1)^{-1} & 0 \\ 0 & (\hat{V}'\hat{V})^{-1} \end{bmatrix}, \\ (\hat{X}_1'\hat{X}_1)^{-1} &= \begin{bmatrix} \hat{Y}'\hat{Y} & \hat{Y}'Z_1 \\ Z_1'\hat{Y} & Z_1'Z_1 \end{bmatrix}^{-1} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \end{aligned}$$

where $M_{11} = \left[(\hat{Y}'\hat{Y}) - \hat{Y}'Z_1(Z_1'Z_1)^{-1}Z_1'\hat{Y}\right]^{-1} = \left[\hat{Y}'M_1\hat{Y}\right]^{-1} = [Y'(M_1 - M)Y]^{-1}$. So,

$$\begin{aligned} (\hat{X}'\hat{X})^{-1}R_*' &= \begin{bmatrix} M_{11} & M_{12} & 0 \\ M_{21} & M_{22} & 0 \\ 0 & 0 & (\hat{V}'\hat{V})^{-1} \end{bmatrix} \begin{bmatrix} I_G \\ 0 \\ -I_G \end{bmatrix} = \begin{bmatrix} M_{11} \\ M_{21} \\ -(\hat{V}'\hat{V})^{-1} \end{bmatrix} \\ R_*(\hat{X}'\hat{X})^{-1}R_*' &= M_{11} + (\hat{V}'\hat{V})^{-1} \\ \hat{\theta}_{*0} - \hat{\theta}_* &= \begin{bmatrix} \hat{\beta} - \tilde{\beta} \\ \hat{\gamma} - \tilde{\gamma} \\ \hat{\beta} - \tilde{b} \end{bmatrix} = \begin{bmatrix} M_{11} \\ M_{21} \\ -(\hat{V}'\hat{V})^{-1} \end{bmatrix} \left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1}(\tilde{b} - \tilde{\beta}). \end{aligned}$$

Hence, we get

$$\hat{\beta} - \tilde{\beta} = M_{11}\left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1}(\tilde{b} - \tilde{\beta}) = M_{11}\left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1}\tilde{a}, \qquad \text{(A.10)}$$

where $\tilde{a} = \tilde{b} - \tilde{\beta}$ is the OLS estimate of a from (3.3). We see from (A.10) that

$$\begin{aligned} \tilde{a} = \tilde{b} - \tilde{\beta} &= \left[M_{11} + (\hat{V}'\hat{V})^{-1}\right] M_{11}^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= \left\{[Y'(M_1 - M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\}[Y'(M_1 - M)Y](\hat{\beta} - \tilde{\beta}). \qquad \text{(A.11)} \end{aligned}$$

So, we have

$$S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) = (R_*\hat{\theta}_*)' \left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*)$$

$$
\begin{aligned}
&= (\tilde{b}-\tilde{\beta})'\left\{[Y'(M_1-M)Y]^{-1}+(\hat{V}'\hat{V})^{-1}\right\}^{-1}(\tilde{b}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'[Y'(M_1-M)Y]\left\{[Y'(M_1-M)Y]^{-1}+(\hat{V}'\hat{V})^{-1}\right\}\times \\
&\quad [Y'(M_1-M)Y](\hat{\beta}-\tilde{\beta}) = (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11}+(Y'MY)^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11}+(Y'M_1Y-M_{11}^{-1})^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta})\,. \qquad \text{(A.12)}
\end{aligned}
$$

Now, we can apply the following lemma which proof is straightforward and then, is omitted.

Lemma A.1 *Let A and B be two nonsingular $r\times r$ matrices. Then*

$$
\begin{aligned}
A^{-1}-B^{-1} &= B^{-1}(B-A)A^{-1} \\
&= A^{-1}(B-A)B^{-1} \\
&= A^{-1}(A-AB^{-1}A)A^{-1} \\
&= B^{-1}(BA^{-1}B-B)B^{-1}.
\end{aligned}
$$

Furthermore, if $B-A$ is nonsingular, then $A^{-1}-B^{-1}$ is nonsingular with

$$
\begin{aligned}
(A^{-1}-B^{-1})^{-1} &= A(B-A)^{-1}B = A+A(B-A)^{-1}A = A[A^{-1}+(B-A)^{-1}]A \\
&= B(B-A)^{-1}A = B(B-A)^{-1}B-B = B[(B-A)^{-1}-B^{-1}]B \\
&= A(A-AB^{-1}A)^{-1}A \\
&= B(BA^{-1}B-B)^{-1}B.
\end{aligned}
$$

By setting $A=M_{11}^{-1}$ and $B=Y'M_1Y$ in (A.12), and applying Lemma **A.1**, we get

$$
\begin{aligned}
S(\hat{\theta}_{*0})-S(\hat{\theta}_{*}) &= (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11}+(Y'M_1Y-M_{11}^{-1})^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'A\left[A^{-1}+(B-A)^{-1}\right]A(\hat{\beta}-\tilde{\beta}) = (\hat{\beta}-\tilde{\beta})'(B^{-1}-A^{-1})^{-1}(\hat{\beta}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'\{[Y'(M_1-M)Y]^{-1}-(Y'M_1Y)^{-1}\}^{-1}(\hat{\beta}-\tilde{\beta}) \\
&= \frac{1}{T}(\tilde{\beta}-\hat{\beta})'[\hat{\Omega}_{IV}^{-1}-\hat{\Omega}_{LS}^{-1}]^{-1}(\tilde{\beta}-\hat{\beta}) = \frac{1}{T}(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta})\,, \qquad \text{(A.13)}
\end{aligned}
$$

where $\hat{\Omega}_{IV}=\frac{1}{T}Y'(M_1-M)Y$ and $\hat{\Omega}_{LS}=\frac{1}{T}Y'M_1Y$. Note also that

$$
S(\hat{\theta}_{*0})-S(\hat{\theta}_{*}) = S(\hat{\theta}_0)-S(\hat{\theta}) = \tilde{a}'[\hat{V}'M_X\hat{V}]\tilde{a}\,, \qquad \text{(A.14)}
$$

where $M_X = I-P_X = I-X(X'X)^{-1}X'$, $X=[Y, Z_1, \hat{V}]$. Moreover, from (3.11), we have

$$
S(\hat{\theta}) = T\tilde{\sigma}_2^2,\; S(\hat{\theta}_0)=T\hat{\sigma}^2,\; S_*(\hat{\theta}_*^0)=T\tilde{\sigma}^2\,. \qquad \text{(A.15)}
$$

Hence, except for H_1, the other statistics can be expressed as:

$$
\begin{aligned}
\mathscr{H}_2 &= T[S(\hat{\theta}_0)-S(\hat{\theta})]/S_*(\hat{\theta}_*^0),\; \mathscr{H}_3 = T[S(\hat{\theta}_0)-S(\hat{\theta})]/S(\hat{\theta}_0)\,, \qquad \text{(A.16)} \\
\mathscr{T}_1 &= \kappa_1[S(\hat{\theta}_0)-S(\hat{\theta})]/[S_*(\hat{\theta}_*^0)-S_e(\hat{\theta})],\; \mathscr{T}_2 = \kappa_2[S(\hat{\theta}_0)-S(\hat{\theta})]/S(\hat{\theta})\,, \qquad \text{(A.17)}
\end{aligned}
$$

$$\mathcal{T}_3 = \kappa_3[S(\hat{\theta}_0) - S(\hat{\theta})]/S_*(\hat{\theta}^0_*),\ \mathcal{T}_4 = \kappa_4[S(\hat{\theta}_0) - S(\hat{\theta})]/S(\hat{\theta}_0)\,, \tag{A.18}$$
$$\mathcal{RH} = \kappa_R[\bar{S}(\hat{\bar{\theta}}_0) - \bar{S}(\hat{\bar{\theta}})]/\bar{S}(\hat{\bar{\theta}}_0)\,, \tag{A.19}$$

Equations (A.16) - (A.19) are the regression interpretation of DWH and RH statistics.

B. Proofs

PROOF OF LEMMA **6.1** Note first that

$$\tilde{\beta} = \beta + [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)u = \beta + \bar{A}_1 u,$$
$$\bar{A}_1 = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)\,, \tag{B.1}$$
$$\hat{\beta} = \beta + (Y'M_1Y)^{-1}Y'M_1u = \beta + A_1u, \quad A_1 = (Y'M_1Y)^{-1}Y'M_1 \tag{B.2}$$
$$\tilde{\beta} - \hat{\beta} = (\bar{A}_1 - A_1)u,\ (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = u'C_0u\,, \tag{B.3}$$

with $C_0 = (\bar{A}_1 - A_1)'\hat{\Delta}^{-1}(\bar{A}_1 - A_1)$. We also have

$$M_1(y - Y\tilde{\beta}) = \bar{B}_1 u\,, \tag{B.4}$$
$$M(y - Y\tilde{\beta}) = Mu - MY\bar{A}_1u = Mu - MM_1Y\bar{A}_1u = MM_{(M_1-M)Y}u\,, \tag{B.5}$$

where $\bar{B}_1 = M_1 - P_{(M_1-M)Y} = M_1(I - P_{(M_1-M)Y}) = M_1M_{(M_1-M)Y}$, and

$$\tilde{\sigma}^2 = \frac{1}{T}u'M_1M_{(M_1-M)Y}u = u'\bar{D}_1u, \quad \hat{\sigma}^2 = \frac{1}{T}u'M_1M_{M_1Y}u = u'D_1u\,, \tag{B.6}$$
$$\tilde{\sigma}_1^2 = \tilde{\sigma}^2 - \hat{\sigma}^2 = u'(\bar{D}_1 - D_1)u = \frac{1}{T}u'(M_1 - M)M_{(M_1-M)Y}u\,, \tag{B.7}$$
$$\tilde{\sigma}_2^2 = \frac{1}{T}u'M_1M_{M_1Y}u - u'C_0u = u'(D_1 - C_0)u\,. \tag{B.8}$$

Now, from (B.1) - (B.8) and the definitions of the statistics, we get:

$$\mathcal{H}_2 = Tu'C_0u/u'\bar{D}_1u = T(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'\bar{D}_1(u/\sigma_u)\,, \tag{B.9}$$
$$\mathcal{H}_3 = Tu'C_0u/u'D_1u = T(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'D_1(u/\sigma_u)\,, \tag{B.10}$$
$$\mathcal{T}_1 = \kappa_1(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'(\bar{D}_1 - D_1)(u/\sigma_u), \tag{B.11}$$
$$\mathcal{T}_2 = \kappa_2(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'(D_1 - C_0)(u/\sigma_u)\,, \tag{B.12}$$
$$\mathcal{T}_3 = \kappa_3(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'\bar{D}_1(u/\sigma_u), \tag{B.13}$$
$$\mathcal{T}_4 = \kappa_4(u/\sigma_u)'C_0(u/\sigma_u)/(u/\sigma_u)'D_1(u/\sigma_u)\,. \tag{B.14}$$

Under H_0, Y is independent of u, and if further the instruments Z are exogenous, the conditional distribution, given $\bar{X}$ of all statistics in (B.9) - (B.14) depend only on the distribution of u/σ_u, irrespective of whether identification is strong or weak. The same result holds for $\mathcal{H}_1$. By observing that $\frac{1}{T}(M_{X_1} - M_{\bar{X}}) = P_{D_1 Z_2}$, $\mathscr{RH}$ can also be expressed as:

$$\mathscr{RH} \quad = \quad \kappa_R (u/\sigma_u)' P_{D_1 Z_2} (u/\sigma_u)/k_2/(u/\sigma_u)'(D_1 - P_{D_1 Z_2})(u/\sigma_u) . \tag{B.15}$$

Thus, under H_0, the distribution of $\mathscr{RH}$, given $\bar{X}$, only depends on u/σ_u, whether $\mathrm{Rank}(\Pi_2) = G$ or not. □

PROOF OF LEMMA **6.1** Consider the identities expressing $\mathscr{H}_i, i = 1, 2, 3,$ $\mathscr{T}_l, l = 1, 2, 3, 4,$ and $\mathscr{RH}$ in (B.9) - (B.15). Under H_1, we have $u = Va + \varepsilon$ and the results of Theorem **5.2** follow. □

PROOF OF LEMMA **6.1** Suppose that $a \in \mathscr{N}(C_\pi)$. Then, we can show that

$$(\bar{A}_1 - A_1)Va \quad = \quad 0, C_0 Va = 0, \bar{D}_1 Va = 0, D_1 Va = 0, \tag{B.16}$$

$$M_{X_1} Va \quad = \quad D_1 Va = 0, M_{\bar{X}} Va = D_1 Va - P_{D_1 Z_2} Va = 0 , \tag{B.17}$$

where $\bar{A}_1$, A_1, C_0, $\bar{D}_1$, and D_1 are defined in (B.1) - (B.8) and (B.15).

To simplify, let us prove that $(\bar{A}_1 - A_1)Va = 0$. First, note that $V = Y - Z_1\Pi_1 - Z_2\Pi_2$ so that $(\bar{A}_1 - A_1)Va = [\bar{A}_1 Y - A_1 Y - (\bar{A}_1 - A_1)(Z_1\Pi_1 + Z_2\Pi_2)]a$. Since $\bar{A}_1 Y = I_G = A_1 Y$, hence we have

$$(\bar{A}_1 - A_1)Va = -[(\bar{A}_1 - A_1)(Z_1\Pi_1 + Z_2\Pi_2)]a = -(\bar{A}_1 - A_1)Z_2\Pi_2 a , \tag{B.18}$$

because $\bar{A}_1 Z_1 = A_1 Z_1 = 0$. Now, we observe that $(M_1 - M)Z_2 = M_1 Z_2$, hence $(\bar{A}_1 - A_1)Z_2\Pi_2 a = (\hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1})M_1 Z_2 \Pi_2 a$, which equals zero if and only $M_1 Z_2 \Pi_2 a = 0$, *i.e.* $\Pi_2' Z_2' M_1 Z_2 \Pi_2 a$ or equivalently, $a \in \mathscr{N}(C_\pi)$. So, we have $a \in \mathscr{N}(C_\pi)$ if and only if $(\bar{A}_1 - A_1)Va = 0$. The proof is similar for the other identities in (B.16)-(B.17). Thus by substituting these identities in Theorem **5.2**, we get the results of Corollary **5.3**.
Suppose now that (5.13)-(5.17) hold. It is easy to see from Theorem **5.2** that this equivalent to

$$(\bar{A}_1 - A_1)Va = 0, C_0 Va = 0, \bar{D}_1 Va = 0, D_1 Va = 0, P_{D_1 Z_2} Va = 0 \tag{B.19}$$

with probability 1. However, we know that (B.19) holds if and only if $a \in \mathscr{N}(C_\pi)$. Hence the result follows. □

PROOF OF LEMMA **6.1** To simplify the proof, let us focus on $\mathscr{H}_3$. We recall that

$$\mathscr{H}_3 = T(\tilde{\beta} - \hat{\beta})' \hat{\Sigma}_3^{-1} (\tilde{\beta} - \hat{\beta}) , \tag{B.20}$$

where $\hat{\beta} = (Y'M_1Y)^{-1}Y'M_1y$, $\tilde{\beta} = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)y$, $\hat{\Sigma}_3 = \hat{\sigma}^2[(Y'(M_1 - M)Y/T)^{-1} - (Y'M_1Y/T)^{-1}]$, and $\hat{\sigma}^2 = (y - Y\hat{\beta})'M_1(y - Y\hat{\beta})/T$. Let us replace y and Y by $y_* = yR_{11} + YR_{21}$ and $Y_* = YR_{22}$ in (B.20). Then, we get:

$$\mathscr{H}_{3*} = T(\tilde{\beta}_* - \hat{\beta}_*)' \hat{\Sigma}_{*3}^{-1} (\tilde{\beta}_* - \hat{\beta}_*) \tag{B.21}$$

where $\hat{\beta}_*,\ \tilde{\beta}_*,\ \hat{\Sigma}_{3*},$ and $\hat{\sigma}_*^2$ are also obtained by replacing y by $y_* = yR_{11} + YR_{21}$ and Y by $Y_* = YR_{22}$. Now, we have:

$$Y_*'M_1Y_* = R_{22}'Y'M_1YR_{22} = R_{22}'Y'M_1YR_{22},\ Y_*'M_1y_* = R_{22}'(Y'M_1yR_{11} + Y'M_1YR_{21}) \tag{B.22}$$

so that we get:

$$\begin{aligned} \hat{\beta}_* &= R_{22}^{-1}(Y'M_1Y)^{-1}(R_{22}^{-1})'R_{22}'(Y'M_1yR_{11} + Y'M_1YR_{21}) = R_{22}^{-1}(\hat{\beta}R_{11} + R_{21})\,, \\ \tilde{\beta}_* &= (Y_*'(M_1 - M)Y_*)^{-1}Y_*'(M_1 - M)y_* = R_{22}^{-1}(\tilde{\beta}R_{11} + R_{21}),\ \tilde{\beta}_* - \hat{\beta}_* = R_{22}^{-1}(\tilde{\beta} - \hat{\beta})R_{11}\,. \end{aligned}$$

Furthermore, we also have

$$(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*'M_1Y_*/T)^{-1} = R_{22}^{-1}\left[(Y'(M_1 - M)Y/T)^{-1} - (Y'M_1Y/T)^{-1}\right](R_{22}^{-1})',$$

and, since $R_{11} > 0$, we get

$$\begin{aligned} &(\tilde{\beta}_* - \hat{\beta}_*)'\left[(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*'M_1Y_*/T)^{-1}\right]^{-1}(\tilde{\beta}_* - \hat{\beta}_*) \\ &= R_{11}^2(\tilde{\beta} - \hat{\beta})'\left[(Y'(M_1 - M)Y/T)^{-1} - (Y'M_1Y/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta})\,. \end{aligned}$$

By the same way, we find

$$\begin{aligned} y_* - \bar{Y}\hat{\beta}_* &= yR_{11} + YR_{22} - YR_{22}\left[R_{22}Y'(M_1 - M)YR_{22}\right]^{-1}YR_{22}'M_1(yR_{11} + YR_{22}) \\ &= yR_{11} + YR_{22} - Y\hat{\beta}R_{11} - YR_{22} = (y - Y\hat{\beta})R_{11}\,. \\ \hat{\sigma}_*^2 &= (y_* - \bar{Y}\hat{\beta}_*)'M_1(y_* - \bar{Y}\hat{\beta}_*)/T = R_{11}^2(y - Y\hat{\beta})'M_1(y - \bar{Y}\hat{\beta})/T = R_{11}^2\hat{\sigma}^2. \end{aligned}$$

Hence, from (B.21), we can see that

$$\begin{aligned} \mathscr{H}_{3*} &= TR_{11}^2(\tilde{\beta} - \hat{\beta})'\left[(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*'M_1Y_*/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta})/R_{11}^2\hat{\sigma}^2 \\ &= T(\tilde{\beta} - \hat{\beta})'\left[\hat{\sigma}^2(Y_*'(M_1 - M)Y_*/T)^{-1} - \hat{\sigma}^2(Y_*'M_1Y_*/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta}) \\ &= \mathscr{H}_3 \end{aligned} \tag{B.23}$$

and the same invariance holds for the author statistics so that Lemma **6.1** follows. □

PROOF OF THEOREM **6.2** Let us replace y by $\bar{y}$ and Y by $\bar{Y}$ in the expressions of the statistics. By Lemma **6.1**, we can write:

$$\mathscr{H}_i = T(\tilde{\beta}_* - \hat{\beta}_*)'\hat{\blacksquare}_{i*}^{-1}(\tilde{\beta}_* - \hat{\beta}_*)\,,\ i = 1, 2, 3\,, \tag{B.24}$$

$$\mathscr{T}_l = \kappa_l(\tilde{\beta}_* - \hat{\beta}_*)'\tilde{\blacksquare}_{l*}^{-1}(\tilde{\beta}_* - \hat{\beta}_*)\,,\ l = 1, 2, 3, 4\,, \tag{B.25}$$

$$\mathscr{R}\mathscr{H} = \kappa_R\vec{y}'\hat{\blacksquare}_{*R}\bar{y}, \tag{B.26}$$

where $\hat{\beta}_*,\ \tilde{\beta}_*,\ \hat{\Sigma}_{*i},\ \tilde{\Sigma}_{*l}$ and $\hat{\Sigma}_{*R}$ are the correspondents of $\hat{\beta},\ \tilde{\beta},\ \hat{\Sigma}_i$ and $\tilde{\Sigma}_l$ defined in (4.2)-(3.10).

From (6.40) and by observing that and $MZ_2 = 0$, we have

$$M\bar{y} = M\bar{v}, M\bar{Y} = M\bar{V}, M_1\bar{y} = M_1(\mu_1 + \bar{v}), M_1\bar{Y} = M_1(\mu_2 + \bar{V}), \tag{B.27}$$

where $\mu_1 = M_1 Z_2 \Pi_2 \zeta = \mu_2 P_{22}^{-1}\zeta$ and $\mu_2 = M_1 Z_2 \Pi_2 P_{22}$, where $\zeta = \beta P_{11} + P_{21}$. From (B.27), we get:

$$\bar{Y}'(M_1 - M)\bar{y} = (\mu_2 + \bar{V})'(M_1 - M)(\mu_1 + \bar{v}), \bar{Y}'M_1\bar{y} = (\mu_2 + \bar{V})'M_1(\mu_1 + \bar{v}), \tag{B.28}$$

$$\bar{Y}'M_1\bar{Y} = (\mu_2 + \bar{V})'M_1(\mu_2 + \bar{V}) = \Omega_{LS}(\mu_2, \bar{V}), \tag{B.29}$$

$$\bar{Y}'(M_1 - M)\bar{Y} = (\mu_2 + \bar{V})'(M_1 - M)(\mu_2 + \bar{V}) = \Omega_{IV}(\mu_2, \bar{V}), \tag{B.30}$$

so that $\hat{\beta}_* = \Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'M_1(\mu_1 + \bar{v})$, $\tilde{\beta}_* = \Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M)(\mu_1 + \bar{v})$, and $\tilde{\beta}_* - \hat{\beta}_* = C(\mu_1 + \bar{v})$, where $C = \Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M) - \Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'M_1$. Moreover, we have $\hat{\sigma}_*^2 = \frac{1}{T}(\bar{y} - \bar{Y}\hat{\beta}_*)'M_1(\bar{y} - \bar{Y}\hat{\beta}_*) = \frac{1}{T}(\mu_1 + \bar{v})'C_*(\mu_1 + \bar{v}) = \frac{1}{T}\omega_{LS}^2(\mu_1, \mu_2, \bar{V}, \bar{v})$, $\tilde{\sigma}_*^2 = \frac{1}{T}(\bar{y} - \bar{Y}\tilde{\beta}_*)'M_1(\bar{y} - \bar{Y}\tilde{\beta}_*) = \frac{1}{T}(\mu_1 + \bar{v})'D_*'D_*(\mu_1 + \bar{v}) = \frac{1}{T}\omega_{IV}^2(\mu_1, \mu_2, \bar{V}, \bar{v})$, with $C_* = \left[I - M_1(\mu_2 + \bar{V})\Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'\right]M_1$ and $D_* = \left[I - M_1(\mu_2 + \bar{V})\Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M)\right]M_1$. Hence, we get

$$\begin{aligned} \hat{\blacksquare}_{1*} &= \omega_{IV}^2(\mu_1, \mu_2, \bar{V}, \bar{v})\Omega_{IV}(\mu_2, \bar{V})^{-1} - \omega_{LS}^2(\mu_1, \mu_2, \bar{V}, \bar{v})\Omega_{LS}(\mu_2, \bar{V})^{-1}, \\ \hat{\blacksquare}_{2*} &= \frac{1}{T}\omega_{IV}^2(\mu_1, \mu_2, \bar{V}, \bar{v})\Delta, \ \hat{\blacksquare}_{3*} = \frac{1}{T}\omega_{LS}^2(\mu_1, \mu_2, \bar{V}, \bar{v})\Delta, \end{aligned} \tag{B.31}$$

where $\Delta = C'C = \Omega_{IV}(\mu_2, \bar{V})^{-1} - \Omega_{LS}(\mu_2, \bar{V})^{-1}$. If $T - k_1 - k_2 > G$, then $\Delta > 0$, thus

$$\mathscr{H}_i = T[\mu_1 + \bar{v}]'\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}], \quad i = 1, 2, 3.$$

where $\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})$, $i = 1, 2, 3$ are defined in Theorem **6.2**. Since $\mathscr{T}_4 = (\kappa_4/T)\mathscr{H}_3$, we find

$$\mathscr{T}_4 = \kappa_4[\mu_1 + \bar{v}]'\Gamma_3(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}]. \tag{B.32}$$

In addition, $\tilde{\sigma}_{*2}^2 = \hat{\sigma}_*^2 - \tilde{\sigma}_*^2(\bar{\beta}_* - \tilde{\beta}_*)'(\bar{\Gamma}_2)^{-1}(\bar{\beta}_* - \tilde{\beta}_*)$ and $\tilde{\sigma}_{*2}^2 = \omega_{LS}^2(\mu_1, \mu_2, \bar{V}, \bar{v}) - (\mu_1 + \bar{v})'C'\Delta^{-1}C(\mu_1 + \bar{v}) = (\mu_1 + \bar{v})'(C_* - C'\Delta^{-1}C)(\mu_1 + \bar{v}) = \omega_2^2(\mu_1, \mu_2, \bar{V}, \bar{v}) \equiv \omega_2^2$, hence, we find

$$\mathscr{T}_2 = \frac{\kappa_2}{\omega_2^2}[\mu_1 + \bar{v}]'C'\Delta^{-1}C[\mu_1 + \bar{v}]. \tag{B.33}$$

In the same way, we also get:

$$\mathscr{T}_l = \frac{\kappa_l}{\omega_l^2}[\mu_1 + \bar{v}]'C'\Delta^{-1}C[\mu_1 + \bar{v}], \quad l = 1, 3, \mathscr{R}\mathscr{H} = \frac{\kappa_R}{\omega_R^2}[\mu_1 + \bar{v}]'P_{D1\bar{Z}_2}[\mu_1 + \bar{v}],$$

where ω_l^2, $l = 1, 3$ and ω_R^2 are defined in Section 3. □

PROOF OF LEMMA **6.1** Set $\Pi_2 a = 0$ in the above proof of Theorem **6.2** and Corollary **6.3** follows.

□

PROOF OF THEOREM **6.4** From Theorem **6.2**, we have

$$\begin{aligned}\mathscr{T}_l &= \kappa_l[\mu_1+\bar{v}]'\bar{\Gamma}_l(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}],\ \mathscr{H}_i = T[\mu_1+\bar{v}]'\Gamma_i(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}],\\ \mathscr{R}\mathscr{H} &= \kappa_R[\mu_1+\bar{v}]'\Gamma_R(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}],\end{aligned}$$

for all $l=1,2,3,4$ and all $i=1,2,3,$ where $\bar{\Gamma}_l(\mu_1,\mu_2,\bar{V},\bar{v}),\ \bar{\Gamma}_i(\mu_1,\mu_2,\bar{V},\bar{v}),\ \Gamma_R(\mu_1,\mu_2,\bar{V},\bar{v}),$ $\mu_1,\ \mu_2,\ \kappa_l$ and κ_R are defined in Section 3.

Assume that Z is fixed. Under the normality assumption (6.27), $\mu_1+\bar{v}$ is independent of $\bar{V}$ and $\mu_1+\bar{v}|_Z \sim \mathrm{N}(\mu_1,1)$. Since $C'\Delta^{-1}C,$ is symmetric idempotent of rank $G,$ C and Δ are defined in Theorem **6.2**, we have $(\mu_1+\bar{v})'C'\Delta^{-1}C(\mu_1+\bar{v})|\bar{V}\sim\chi^2(G,\nu_1),$ where $\nu_1=\mu_1'C'\Delta^{-1}C\mu_1.$ By the same way, the denominator of $\mathscr{T}_1$ (without the scaling factor) is $(\mu_1+\bar{v})'E(\mu_1+\bar{v})|\bar{V}\sim\chi^2(k_2-G,\upsilon_1),$ where E defined in Section 3 is symmetric idempotent of rank $k_2-G,$ and with $\upsilon_1=\mu_1'E\mu_1.$ Furthermore, we have $(C'\Delta^{-1}C)E=0,$ hence

$$\mathscr{T}_1|\bar{V}\sim F(G,k_2-G;\nu_1,\upsilon_1). \tag{B.34}$$

By the same way, we get:

$$\mathscr{T}_2|\bar{V}\sim F(G,T-k_1-2G;\nu_1,\upsilon_2), \tag{B.35}$$

where $\upsilon_2=\mu_1'(C_*-C'\Delta^{-1}C)\mu_1.$ Now, from the notations in Theorem **6.2**, we can write:

$$\mathscr{T}_4=\kappa_4/(1+\frac{1}{\kappa_2\mathscr{T}_2}), \tag{B.36}$$

and since $\mathscr{T}_2|\bar{V}\sim F(G,T-k_1-2G;\nu_1,\upsilon_2),$ we have $\frac{1}{\mathscr{T}_2}|\bar{V}\sim F(T-k_1-2G,G;\upsilon_2,\nu_1)$ so that

$$\mathscr{T}_4|\bar{V}\sim\kappa_4/[1+\frac{1}{\kappa_2}F(T-k_1-2G,G;\upsilon_2,\nu_1)]. \tag{B.37}$$

Note also that $\omega_{LS}^2\geq\omega_2^2$ entails that

$$\mathscr{T}_4|\bar{V}\leq\frac{\kappa_4}{\omega_2^2}(\mu_1+\bar{v})'C'\Delta^{-1}C(\mu_1+\bar{v})|\bar{V}=\bar{\kappa}_2^*\mathscr{T}_2|\bar{V}\sim\bar{\kappa}_2^*F(G,T-k_1-2G;\nu_1,\upsilon_2), \tag{B.38}$$

where $\kappa_2,\ \kappa_4,\ \bar{\kappa}_2^*$ are given in Theorem **6.4**. For $\mathscr{T}_3$, we note that its numerator and denominator are such that

$$\begin{aligned}(\mu_1+\bar{v})'C'\Delta^{-1}C(\mu_1+\bar{v})|\bar{V}\sim\chi^2(G;\nu_1),\ \omega_{IV}^2 &= (\mu_1+\bar{v})'D_*'D_*(\mu_1+\bar{v})\\ &\sim \chi^2(T-k_1-G;\nu_3),\end{aligned} \tag{B.39}$$

where $\nu_3=\mu_1'D_*'D_*\mu_1.$ Since $D_*'D_*(C'\Delta^{-1}C)\neq0,$ $\mathscr{T}_3$ does not follow necessary a F-distribution.

By the same way, we get the results for $\mathscr{H}_2$, $\mathscr{H}_3$ and $\mathscr{R}\mathscr{H}$. □

References

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Bekker, P., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

Cragg, J., Donald, S., 1993. Testing identifiability and specification in instrumental variables models. Econometric Theory 9, 222–240.

Doko, F., Dufour, J.-M., 2008. Instrument endogeneity and identification-robust tests: some analytical results. Journal of Statistical Planning and Inference 138(9), 2649–2661.

Doko, T. F. , Dufour, J.-M. , 2011. Exogeneity tests and estimation in IV regressions. Technical report, Department of Economics, McGill University, Canada Montréal, Canada.

Donald, S. G., Newey, W. K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.

Dufour, J.-M., 1987. Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In: I. B. MacNeill , G. J. Umphrey, eds, Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling. D. Reidel, Dordrecht, The Netherlands, pp. 317–338.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M. , 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36(4), 767–808.

Dufour, J.-M. , 2006. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 138, 2649–2661.

Dufour, J.-M. , Khalaf, L. , 2002. Simulation based finite and large sample tests in multivariate regressions. Journal of Econometrics 111(2), 303–322.

Dufour, J.-M., Khalaf, L., Beaulieu, M.-C., 2010. Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models. Journal of Applied Econometrics 25, 263–285.

Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73(4), 1351–1365.

Dufour, J.-M., Taamouti, M., 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Guggenberger, P., 2010. The impact of a Hausman pretest on the size of the hypothesis tests. Econometric Theory 156, 337–343.

Hahn, J., Ham, J., Moon, H. R., 2010. The Hausman test and weak instruments. Journal of Econometrics 160, 289–299.

Hall, A. R., Peixe, F. P. M., 2003. A consistent method for the selection of relevant instruments. Econometric Reviews 2(3), 269–287.

Hall, A. R., Rudebusch, G. D., Wilcox, D. W., 1996. Judging instrument relevance in instrumental variables estimation. International Economic Review 37, 283–298.

Hansen, L. P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L. P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Hansen, L. P., Singleton, K., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50, 1269–1286.

Harville, D. A., 1997. Matrix Algebra from a Statistician's Perspective. Springer-Verlag, New York.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70(5), 1781–1803.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1124.

Moreira, M. J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71(4), 1027–1048.

Nelson, C., Startz, R., 1990a. The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one. The Journal of Business 63, 125–140.

Nelson, C., Startz, R., 1990b. Some further results on the exact small properties of the instrumental variable estimator. Econometrica 58, 967–976.

Phillips, P. C. B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

Rao, T., Mitra, C., eds, 1971. Generalized Inverse of Matrices and its Applications. John Wiley & Sons, Inc. Hoboken, New Jersey.

Revankar, N. S., Hartley, M. J., 1973. An independence test and conditional unbiased predictions in the context of simultaneous equation systems. International Economic Review 14, 625–631.

Sargan, J., 1983. Identification and lack of identification,. Econometrica 51, 1605–1633.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65(3), 557–586.

Stock, J. H., Wright, J. H., 2000. GMM with weak identification. Econometrica 68, 1055–1096.

Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20(4), 518–529.

Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66(6), 1389–1404.

Wu, D.-M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41, 733–750.