Exogeneity tests, incomplete models, weak identification and non-Gaussian distributions: invariance and finite-sample distributional theory *

Firmin Doko Tchatoka† — The University of Adelaide

Jean-Marie Dufour ‡ — McGill University

First version: May 2007
Revised: October 2007, November 2008, December 2009, March 2010, January 2012, July 2013, September 2015, December 2015, May 2016, December 2016, September 2017, January 2018, July 2018
This version: July 2018
Compiled: May 15, 2020, 22:29

* The authors thank Nazmul Ahsan, Marine Carrasco, Atsushi Inoue, Jan Kiviet, Vinh Nguyen, Benoit Perron, Pascale Valéry, Hui Jun Zhang, the Editors – Marine Carrasco, Marcelo Moreira, Benoit Perron, and Victoria Zinde-Walsh – and two anonymous referees for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), 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, and the Fonds de recherche sur la société et la culture (Québec).

† School of Economics, The University of Adelaide, 10 Pulteney Street, Adelaide, SA 5005. Tel: +618 8313 1174; e-mail: firmin.dokotchatoka@adelaide.edu.au. Homepage: http://www.adelaide.edu.au/directory/firmin.dokotchatoka

‡ 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 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca`. Web page: `http://www.jeanmariedufour.com`

ABSTRACT

We study the distribution of Durbin-Wu-Hausman (DWH) and Revankar-Hartley (RH) tests for exogeneity from a finite-sample viewpoint, under the null and alternative hypotheses. We consider linear structural models with possibly non-Gaussian errors, where structural parameters may not be identified and where reduced forms can be incompletely specified (or nonparametric). On level control, we characterize the null distributions of all the test statistics. Through conditioning and invariance arguments, we show that these distributions do not involve nuisance parameters. In particular, this applies to several test statistics for which no finite-sample distributional theory is yet available, such as the standard statistic proposed by Hausman (1978). The distributions of the test statistics may be non-standard – so corrections to usual asymptotic critical values are needed – but the characterizations are sufficiently explicit to yield finite-sample (Monte-Carlo) tests of the exogeneity hypothesis. The procedures so obtained are robust to weak identification, missing instruments or misspecified reduced forms, and can easily be adapted to allow for parametric non-Gaussian error distributions. We give a general invariance result (*block triangular invariance*) for exogeneity test statistics. This property yields a convenient *exogeneity canonical form* and a parsimonious reduction of the parameters on which power depends. In the extreme case where no structural parameter is identified, the distributions under the alternative hypothesis and the null hypothesis are identical, so the power function is flat, for all the exogeneity statistics. However, as soon as identification does not fail completely, this phenomenon typically disappears. We present simulation evidence which confirms the finite-sample theory. The theoretical results are illustrated with two empirical examples : the relation between trade and economic growth, and the widely studied problem of the return of education to earnings.

Keywords : Exogeneity; Durbin-Wu-Hausman test; weak instrument; incomplete model; non-Gaussian; weak identification; identification robust; finite-sample theory; pivotal; invariance; Monte Carlo test; power.

JEL classification : C3; C12; C15; C52.

Contents

List of Tables

List of Definitions, Assumptions, Propositions and Theorems

1. Introduction

The literature on weak instruments is now considerable and has often focused on inference for the coefficients of endogenous variables in so-called "instrumental-variable regressions" (or "IV regressions"); see the reviews of Stock, Wright and Yogo (2002), Dufour (2003), Andrews and Stock (2007), and Poskitt and Skeels (2012). Although research on tests for exogeneity in IV regressions is considerable, most of these studies either deal with cases where instrumental variables are strong (thus leaving out issues related to weak instruments), or focus on the asymptotic properties of exogeneity tests.[1] To the best of our knowledge, there is no study on the finite-sample performance of exogeneity tests when IVs can be arbitrary weak, when the errors may follow a non-Gaussian distribution, or when the reduced form is incompletely specified. The latter feature is especially important to avoid losing the validity of the test procedure when important instruments are "left-out" when applying an exogeneity test, as happens easily for some common "identification-robust" tests on model structural coefficients [see Dufour and Taamouti (2007)].

In this paper, we investigate the finite-sample properties (size and power) of exogeneity tests of the type proposed by Durbin (1954), Wu (1973), Hausman (1978), and Revankar and Hartley (1973), henceforth DWH and RH tests, allowing for: (a) the possibility of identification failure (weak instruments); (b) model errors with non-Gaussian distributions, including heavy-tailed distributions which may lack moments (such as the Cauchy distribution); and (c) incomplete reduced forms (*e.g.*, situations where important instruments are missing or left out) and arbitrary heterogeneity in the reduced forms of potentially endogenous explanatory variables.

As pointed out early by Wu (1973), a number of economic hypotheses can be formulated in terms of independence (or "exogeneity") between stochastic explanatory variables and the disturbance term in an equation. These include, for example, the permanent income hypothesis, expected profit maximization, and recursiveness hypotheses in simultaneous equations. Exogeneity (or "predetermination") assumptions can also affect the "causal interpretation" of model coefficients [see Simon (1953), Engle, Hendry and Richard (1982), Angrist and Pischke (2009), Pearl (2009)], and eventually the choice of estimation method.

To achieve the above goals, we consider a general setup which allows for non-Gaussian distributions and arbitrary heterogeneity in reduced-form errors. Under the assumption that the distribution of the structural errors (given IVs) is specified up to an unknown factor (which may depend on IVs), we show that exact exogeneity tests can be obtained from all DWH and RH statistics [including the one involving both ordinary-least-squares OLS and IV-based error variance estimates, as proposed by Hausman (1978)] through the Monte Carlo test (MCT) method [see Dufour (2006)]. The null

[1] See, for example, Durbin (1954), Wu (1973, 1974, 1983*a*, 1983*b*), Revankar and Hartley (1973), Farebrother (1976), Hausman (1978), Revankar (1978), Dufour (1979, 1987), Hwang (1980, 1985), Kariya and Hodoshima (1980), Hausman and Taylor (1981), Spencer and Berk (1981), Nakamura and Nakamura (1981, 1985), Engle (1982), Holly (1982, 1983*b*, 1983*a*), Holly and Monfort (1983), Reynolds (1982), Smith (1983, 1984, 1985, 1994), Thurman (1986), Rivers and Vuong (1988), Smith and Pesaran (1990), Ruud (1984, 2000), Newey (1985*a*, 1985*b*), Davidson and Mackinnon (1985, 1985, 1989, 1990, 1993), Meepagala (1992), Wong (1996, 1997), Ahn (1997), Staiger and Stock (1997), Hahn and Hausman (2002), Baum, Schaffer and Stillman (2003), Kiviet and Niemczyk (2006, 2007), Blundell and Horowitz (2007), Guggenberger (2010), Hahn, Ham and Moon (2010), Jeong and Yoon (2010), Chmelarova and Hill (2010), Kiviet and Pleus (2012), Lee and Okui (2012), Kiviet (2013), Wooldridge (2014, 2015), Caetano (2015), Doko Tchatoka (2015*a*), Kabaila, Mainzer and Farchione (2015), and Lochner and Moretti (2015).

distributions of the test statistics typically depend on specific instrument values, so "critical values" should also depend on the latter. Despite this, the MCT procedure automatically controls the level irrespective of this complication, and thus *avoids* the need to compute *critical values*. Of course, as usual, the null hypothesis is interpreted here as the conjunction of all model assumptions (including "distributional" ones) with the exogeneity restriction.

The finite-sample tests built in this way are also robust to weak instruments, in the sense that they never over-reject the null hypothesis of exogeneity even when IVs are weak. This entails that size control is feasible in finite samples for all DWH and RH tests [including the Hausman (1978) test]. All exogeneity tests considered can also be described as identification-robust in finite samples. These conclusions stand in contrast with ones reached by Staiger and Stock (1997, Section D) who argue – following a local asymptotic theory – that size adjustment may not be feasible due to the presence of nuisance parameters in the asymptotic distribution. Of course, this underscores the fundamental difference between a finite-sample theory and an asymptotic approximation, even when the latter is "improved".

More importantly, we show that the proposed MCT procedure remains valid even if the right-hand-side (possibly) endogenous regressors are heterogenous and the reduced-form model is incompletely specified (missing instruments). Because of the latter property, we say that the DWH and RH tests are *robust to incomplete reduced forms*. For example, robustness to incomplete reduced forms is relevant in macroeconomic models with structural breaks in the reduced form : this shows that exogeneity tests remain applicable without knowledge of break dates. In such contexts, inference on the structural form may be more reliable than inference on the reduced form. This is of great practical interest, for example, in inference based on IV regressions and DSGE models. For further discussion of this issue, see Dufour and Taamouti (2007), Kleibergen and Mavroeidis (2009), Dufour, Khalaf and Kichian (2013), and Doko Tchatoka (2015*b*).

We study analytically the power of the tests and identify the crucial parameters of the power function. In order to do this, we first prove a general invariance property (*block triangular invariance*) for exogeneity test statistics – a result of separate interest, *e.g.* to study how nuisance parameters may affect the distributions of exogeneity test statistics. This property yields a convenient *exogeneity canonical form* and a parsimonious reduction of the parameters on which power depends. In particular, we give conditions under which exogeneity tests have no power, and conditions under which they have power. We show formally that the tests have little power when instruments are weak. In particular, the power of the tests cannot exceed the nominal level if all structural parameters are completely unidentified. Nevertheless, power may exist as soon as one instrument is strong (partial identification).

We present a Monte Carlo experiment which confirms our theoretical findings. In particular, simulation results confirm that the MCT versions of all exogeneity statistics considered allow one to control test size perfectly, while usual critical values (under a Gaussian error assumption) are either exact or conservative. The conservative property is visible in particular when the two-stage-least-squares (2SLS) estimator of the structural error variance is used in covariance matrices. In such cases, the MCT version of the tests allows sizable power gains. The results are also illustrated through an empirical example : the relation between trade and economic growth.

The paper is organized as follows. Section 2 formulates the model studied, and Section 3 describes the exogeneity test statistics, including a number of alternative formulations (*e.g.*, linear-regression-based interpretations) which may have different analytical and numerical features. In Section 4, we give general characterizations of the finite-sample distributions of the test statistics and show how they can be implemented as Monte Carlo (MC) tests, with either Gaussian or non-Gaussian errors. In Section 5, we give the general block-triangular invariance result and describe the associated exogeneity canonical representation. Power is discussed in Section 6. The simulation experiment is presented in Section 7, and the empirical illustration in Section 8. We conclude in Section 9. Additional details on the formulation of the different test statistics and the proofs are supplied in Appendix.

Throughout the paper, I_m stands for the identity matrix of order m. For any full-column-rank $T\times m$ matrix A, $\bar{\mathrm{P}}[A] = A(A'A)^{-1}A'$ is the projection matrix on the space spanned by the columns of A, and $\bar{\mathrm{M}}[A] = I_T - \bar{\mathrm{P}}[A]$. For arbitrary $m\times m$ matrices A and B, the notation $A > 0$ means that A is positive definite (p.d.), $A \geq 0$ means A is positive semidefinite (p.s.d.), and $A \leq B$ means $B - A \geq 0$. Finally, $\|A\|$ is the Euclidian norm of a vector or matrix, *i.e.*, $\|A\| = [\mathrm{tr}(A'A)]^{\frac{1}{2}}$.

2. Framework

We consider a structural model of the form :

$$y = Y\beta + X_1\gamma + u\,, \tag{2.1}$$

$$Y = g(X_1, X_2, X_3, V, \bar{\Pi})\,, \tag{2.2}$$

where (2.1) is a linear structural equation, $y \in \mathbb{R}^T$ 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 which are determined by equation (2.2), $X_1 \in \mathbb{R}^{T\times k_1}$ is a matrix of observations on exogenous variables included in the structural equation (2.1), $X_2 \in \mathbb{R}^{T\times k_2}$ and $X_3 \in \mathbb{R}^{T\times k_3}$ are matrices of observations on exogenous variables excluded from the structural equation, $u = (u_1, \ldots, u_T)' \in \mathbb{R}^T$ is a vector of structural disturbances, $V = [V_1, \ldots, V_T]' \in \mathbb{R}^{T\times G}$ is a matrix of random disturbances, $\beta \in \mathbb{R}^G$ and $\gamma \in \mathbb{R}^{k_1}$ are vectors of unknown fixed structural coefficients, and $\bar{\Pi}$ is a matrix of fixed (typically unknown) coefficients. We suppose $G \geq 1$, $k_1 \geq 0$, $k_2 \geq G$, $k_3 \geq 0$, and denote :

$$X = [X_1, X_2] = [x_1, \ldots, x_T]'\,, \quad \bar{X} = [X_1, X_2, X_3] = [\bar{x}_1, \ldots, \bar{x}_T]'\,, \tag{2.3}$$

$$\bar{Y} = [Y, X_1]\,, \quad Z = [Y, X_1, X_2] = [z_1, \ldots, z_T]'\,, \quad \bar{Z} = [Y, X_1, X_2, X_3] = [\bar{z}_1, \ldots, \bar{z}_T]'\,, \tag{2.4}$$

$$U = [u, V] = [U_1, \ldots, U_T]'\,. \tag{2.5}$$

Equation (2.2) usually represents a reduced-form equation for Y. The form of the function $g(\cdot)$ may be nonlinear or unspecified, so model (2.2) can be viewed as "nonparametric" or "semiparametric". The inclusion of X_3 in this setup allows for Y to depend on exogenous variables not used by the exogeneity tests. This assumption is crucial, because it characterizes the fact that we consider here "incomplete models" where the reduced form for Y may not be specified and involves unknown exogenous variables. It is well known that several "identification-robust" tests for β [such as those proposed by Kleibergen (2002) and Moreira (2003)] are not robust to allowing a general reduced form for Y such as the one in (2.2); see Dufour and Taamouti (2007) and Doko Tchatoka (2015*b*).

We also make the following rank assumption on the matrices $[Y, X]$ and $\left[\bar{\mathrm{P}}[X]Y, X_1\right]$:

$$[Y, X] \text{ and } \left[\bar{\mathrm{P}}[X]Y, X_1\right] \text{ have full-column rank with probability one (conditional on } X). \tag{2.6}$$

This (fairly standard) condition ensures that the matrices X, $\bar{\mathrm{M}}[X_1]Y$ and $\bar{\mathrm{M}}[X]Y$ have full-column rank, hence the unicity of the least-squares (LS) estimates when each column of Y is regressed on X, as well as the existence of a unique 2SLS estimate for β and γ based on X as the instrument matrix. Clearly, (2.6) holds when X has full-column rank and the conditional distribution of Y given X is absolutely continuous (with respect to the Lebesgue measure).

A common additional maintained hypothesis in this context consists in assuming that $g(\cdot)$ is a linear equation of the form

$$Y = X_1\Pi_1 + X_2\Pi_2 + V = X\Pi + V \tag{2.7}$$

where $\Pi_1 \in \mathbb{R}^{k_1\times G}$ and $\Pi_2 \in \mathbb{R}^{k_2\times G}$ are matrices of unknown reduced-form coefficients. In this case, the reduced form for y is

$$y = X_1\pi_1 + X_2\pi_2 + v \tag{2.8}$$

where $\pi_1 = \gamma + \Pi_1\beta$, $\pi_2 = \Pi_2\beta$, and $v = u + V\beta$. When the errors u and V have mean zero (though this assumption may also be replaced by another "location assumption", such as zero medians), the usual necessary and sufficient condition for the identification of this model is

$$\operatorname{rank}(\Pi_2) = G. \tag{2.9}$$

If $\Pi_2 = 0$, the instruments X_2 are irrelevant, and β is completely unidentified. If $1 \le \operatorname{rank}(\Pi_2) < G$, β is not identifiable, but some linear combinations of the elements of β are identifiable [see Dufour and Hsiao (2008) and Doko Tchatoka (2015*b*)]. If Π_2 is close not to have full-column rank [*e.g.*, if some eigenvalues of $\Pi_2'\Pi_2$ are close to zero], some linear combinations of β are ill-determined by the data, a situation often called "weak identification" in this type of setup [see Dufour (2003), Andrews and Stock (2007)].

We study here, from a finite-sample viewpoint, the size and power of the tests proposed by Durbin (1954), Wu (1973), Hausman (1978), and Revankar and Hartley (1973) for assessing the exogeneity of Y in (2.1) - (2.7) when : (a) instruments may be weak; (b) $[u, V]$ may not follow a Gaussian distribution [*e.g.*, heavy-tailed distributions which may lack moments (such as the Cauchy distribution) are allowed]; and (c) the usual reduced-form specification (2.7) is misspecified, and Y follows the more general model (2.2) which allows for omitted instruments, an unspecified nonlinear form and heterogeneity. To achieve this, we consider the following distributional assumptions on model disturbances (where $\mathbb{P}[\cdot]$ refers to the relevant probability measure).

Assumption 2.1 CONDITIONAL SCALE MODEL FOR THE STRUCTURAL ERROR DISTRIBUTION. *For some fixed vector a in $\mathbb{R}^G$, we have :*

$$u = Va + e\,, \tag{2.10}$$

$$e = (e_1, \ldots, e_T)' = \sigma_1(\bar{X})\,\varepsilon\,, \tag{2.11}$$

where $\sigma_1(\bar{X})$ is a (possibly random) function of $\bar{X}$ such that $\mathbb{P}[\sigma_1(\bar{X}) \neq 0 \,|\, \bar{X}] = 1$, and the conditional distribution of ε given $\bar{X}$ is completely specified.

Assumption 2.2 CONDITIONAL MUTUAL INDEPENDENCE OF ε AND V. *V and ε are independent, conditional on $\bar{X}$.*

Assumption 2.3 NONSINGULAR REDUCED-FORM ERRORS. *V has full-column rank with probability one.*

In the above assumptions, possible dependence between u and V is parameterized by a, while ε is independent of V (conditional on $\bar{X}$), and $\sigma_1(\bar{X})$ is an arbitrary (possibly random) *scale parameter* which may depend on $\bar{X}$ (except for the non-degeneracy condition $\mathbb{P}[\sigma_1(\bar{X}) \neq 0 \,|\, \bar{X}] = 1$). So we call a the "endogeneity parameter" of the model. Assumption 2.1 is quite general and allows for heterogeneity in the distributions of the reduced-form disturbances V_t, $t = 1, \dots, T$. In particular, the rows of V need not be identically distributed or independent. Further, non-Gaussian distributions are covered, including heavy-tailed distributions which may lack second moments (such as the Cauchy distribution). In such cases, $\sigma_1(\bar{X})^2$ *does not represent a variance*. Since the scale factor may be random, we can have $\sigma_1(\bar{X}) = \bar{\sigma}(\bar{X}, V, e)$. Of course, these conditions hold when $u = \sigma\varepsilon$, where σ is an unknown positive constant and ε is independent of X with a completely specified distribution. In this context, the standard Gaussian assumption is obtained by taking: $\varepsilon \sim \mathrm{N}[0, I_T]$. Clearly, ε can be multiplied by an arbitrary nonzero constant c, provided $\sigma_1(\bar{X})$ is divided by c. So, when ε has finite second moments and $\mathbb{E}(\varepsilon\varepsilon' \,|\, \bar{X}) = \delta I_T$ with $\delta > 0$, we can set $\delta = 1$ without loss of generality. The distributions of ε and σ_1 may also depend on a subset of $\bar{X}$, such as $X = [X_1, X_2]$. The parameter a is not presumed to be identifiable, and e may not be independent of V – though this would be a reasonable additional assumption to consider in the present context.

We are interested in testing the hypothesis that Y is (strictly) *exogenous* in (2.1). It is clear from (2.2) that this amounts to testing the independence between V and u conditional on $\bar{X}$. In addition, under Assumptions 2.1 and 2.2, $a = 0$ is clearly a sufficient condition for u and V to be independent. Further, as soon as V has full-column rank with probability one (Assumption 2.3), $a = 0$ is also necessary for the latter independence property. This leads to the following equivalent characterization of the (strict) exogeneity of Y with respect to u.

Definition 2.1 STRICT EXOGENEITY. *Y is strictly exogenous with respect to u (conditional on $\bar{X} = [X_1, X_2, X_3]$) in model (2.1)–(2.2) if and only if V and u are independent given $\bar{X}$.*

Under the Assumptions 2.1 - 2.3, it is then easy to see that the exogeneity hypothesis on Y (as defined above) can be expressed as:

$$H_0 : a = 0\,. \tag{2.12}$$

We stress here that "exogeneity" may depend on a set of conditioning variables ($\bar{X}$), though of course we can have cases where it does not depend on $\bar{X}$ or holds unconditionally. The setup we consider in this paper allows for both possibilities. If the conditional mean of u, given Y and $\bar{X}$, is well defined, then the standard condition for the exogeneity of Y with respect to u in (2.1) is usually expressed as $\mathbb{E}[u \,|\, Y, \bar{X}] = 0$. The notion of exogeneity considered in Definition 2.1 is weaker than

$\mathbb{E}[u\,|\,Y,\bar{X}]=0$ because it does not require the existence of moments for the conditional distribution of u given $(Y,\bar{X})$, and stronger because an independence assumption (between V and u) is used. Clearly this notion is not synonym with the with non-correlation This notion of exogeneity allows one to cover models in which the errors have heavy-tailed distributions that may lack moments (such as the Cauchy or $t(1)$ distributions). It is important to note that a notion of "exogeneity" based on second moments becomes empty in the presence of heavy tails, especially if second moments do not exist; for a discussion of the fragility of moment-based hypotheses, see Dufour (2003). The bulk of the literature on weak instruments has focused on testing the parameters β and γ of equation (2.1), rather than a; see Dufour (1997), Dufour and Jasiak (2001), Kleibergen (2002), Moreira (2003), and the reviews of Stock et al. (2002), Dufour (2003), Andrews and Stock (2007), and Poskitt and Skeels (2012).

Clearly, the independence assumption is not synonym with "non-correlation", but it plays a crucial role if one wishes to obtain a finite-sample distributional theory. It is important to be aware that hypotheses and assumptions about moments are not generally testable in nonparametric setups (which are usually considered in asymptotic theory); see Bahadur and Savage (1956) and Dufour (2003). Despite these limitations, it can be illuminating to study how H_0 can be reinterpreted in the familiar language of covariance hypotheses, provided standard second-moment assumptions are made, as follows.

Assumption 2.4 HOMOSKEDASTICITY. *The vectors $U_t=[u_t,V_t']'$, $t=1,\ldots,T$, have zero means and the same (finite) nonsingular covariance matrix:*

$$\mathbb{E}[U_tU_t^{'}\,|\,\bar{X}]=\Sigma=\begin{bmatrix}\sigma_u^2 & \sigma_{Vu}^{'}\\ \sigma_{Vu} & \Sigma_V\end{bmatrix}>0\,,\quad t=1,\ldots,T. \tag{2.13}$$

where σ_u^2, σ_{Vu} and Σ_V may depend on $\bar{X}$.

Assumption 2.5 ORTHOGONALITY BETWEEN e AND V. $\mathbb{E}[V_t\,e_t\,|\,\bar{X}]=0$, $\mathbb{E}[e_t\,|\,\bar{X}]=0$ *and* $\mathbb{E}[e_t^2\,|\,\bar{X}]=\sigma_e^2$, *for $t=1,\ldots,T$.*

Under these two assumptions, the reduced-form disturbances

$$W_t=[v_t,\,V_t']'=[u_t+V_t'\beta,\,V_t']',\quad t=1,\ldots,T, \tag{2.14}$$

also have a nonsingular covariance matrix (conditional on $\bar{X}$),

$$\Omega=\begin{bmatrix}\sigma_u^2+\beta'\Sigma_V\beta+2\beta'\sigma_{Vu} & \beta'\Sigma_V+\sigma_{Vu}'\\ \Sigma_V\beta+\sigma_{Vu} & \Sigma_V\end{bmatrix}. \tag{2.15}$$

In this context, the exogeneity hypothesis of Y can be formulated as

$$H_0:\sigma_{Vu}=0\,. \tag{2.16}$$

Further,

$$\sigma_{Vu}=\Sigma_Va\,,\quad \sigma_u^2=\sigma_e^2+a'\Sigma_Va=\sigma_e^2+\sigma_{Vu}'\Sigma_V^{-1}\sigma_{Vu}\,, \tag{2.17}$$

so $\sigma_{Vu} = 0 \Leftrightarrow a = 0$, and the exogeneity of Y can be assessed by testing whether $a = 0$. Note, however, that Assumptions 2.4 and 2.5 will not be needed for the results presented in this paper.

In order to study the power of exogeneity tests, it will be useful to consider the following separability assumptions.

Assumption 2.6 ENDOGENEITY-PARAMETER DISTRIBUTIONAL SEPARABILITY. *$\bar{\Pi}$ is not restricted by a, and the conditional distribution of $[V, e]$ given $\bar{X}$ does not depend on the parameter a.*

Assumption 2.7 REDUCED-FORM LINEAR SEPARABILITY FOR Y. *Y satisfies the equation*

$$Y = g(X_1, X_2, X_3, \bar{\Pi}) + V\,. \tag{2.18}$$

Assumption 2.6 means that the distributions of V and e do not depend on the endogeneity parameter a, while Assumption 2.7 means that V can be linearly separated from $g(X_1, X_2, X_3, \bar{\Pi})$ in (2.2).

3. Exogeneity tests

We consider the four statistics proposed by Wu (1973) [$\mathscr{T}_l, l = 1, 2, 3, 4$], the statistic proposed by Hausman (1978) [$\mathscr{H}_1$] as well as some variants [$\mathscr{H}_2, \mathscr{H}_3$] occasionally considered in the literature [see Hahn et al. (2010)], and the test suggested by Revankar and Hartley (1973, RH) [$\mathscr{R}$]. These statistics can be formulated in two alternative ways: (1) as Wald-type statistics for the difference between the 2SLS and OLS estimators of β in equation (2.1), where different statistics are obtained by changing the covariance matrix; or (2) a F-type significance test on the coefficients of an "extended" version of (2.1), so the different statistics can be written in terms of the difference between restricted and unrestricted residual sum of squares.

3.1. Test statistics

We now give a unified presentation of different available DWH-type statistics. The test statistics considered can be written as follows:

$$\mathscr{T}_i = \kappa_i(\tilde{\beta} - \hat{\beta})' \tilde{\Sigma}_i^{-1} (\tilde{\beta} - \hat{\beta})\,, \quad i = 1, 2, 3, 4, \tag{3.1}$$

$$\mathscr{H}_j = T(\tilde{\beta} - \hat{\beta})' \hat{\Sigma}_j^{-1} (\tilde{\beta} - \hat{\beta})\,, \quad j = 1, 2, 3, \tag{3.2}$$

$$\mathscr{R} = \kappa_R \left(y' \Psi_R y \,/\, \hat{\sigma}_R^2 \right), \tag{3.3}$$

where $\hat{\beta}$ and $\tilde{\beta}$ are OLS and 2SLS estimators of β, *i.e.*

$$\hat{\beta} = (Y' M_1 Y)^{-1} Y' M_1 y\,, \tag{3.4}$$

$$\tilde{\beta} = [(PY)' M_1 (PY)]^{-1} (PY)' M_1 y = (Y' N_1 Y)^{-1} Y' N_1 y\,, \tag{3.5}$$

while we denote $\hat{\gamma}$ and $\tilde{\gamma}$ the corresponding OLS and 2SLS estimators of γ, and

$$M_1 = \bar{\mathrm{M}}[X_1], \quad P = \bar{\mathrm{P}}[X], \quad M = \bar{\mathrm{M}}[X] = I_T - \bar{\mathrm{P}}[X], \quad N_1 = M_1 P, \tag{3.6}$$

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

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

$$\hat{\Delta} = \hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}, \quad \hat{\Omega}_{IV} = \frac{1}{T} Y' N_1 Y, \quad \hat{\Omega}_{LS} = \frac{1}{T} Y' M_1 Y, \tag{3.9}$$

$$\hat{u} = y - Y\hat{\beta} - X_1\hat{\gamma} = M_1(y - Y\hat{\beta}), \quad \tilde{u} = y - Y\tilde{\beta} - X_1\tilde{\gamma} = M_1(y - Y\tilde{\beta}), \tag{3.10}$$

$$\hat{\sigma}^2 = \frac{1}{T}\hat{u}'\hat{u} = \frac{1}{T}(y - Y\hat{\beta})' M_1 (y - Y\hat{\beta}), \quad \tilde{\sigma}^2 = \frac{1}{T}\tilde{u}'\tilde{u} = \frac{1}{T}(y - Y\tilde{\beta})' M_1 (y - Y\tilde{\beta}), \tag{3.11}$$

$$\tilde{\sigma}_1^2 = \frac{1}{T}(y - Y\tilde{\beta})' N_1 (y - Y\tilde{\beta}) = \tilde{\sigma}^2 - \tilde{\sigma}_e^2, \quad \tilde{\sigma}_e^2 = \frac{1}{T}(y - Y\tilde{\beta})' M (y - Y\tilde{\beta}), \tag{3.12}$$

$$\tilde{\sigma}_2^2 = \hat{\sigma}^2 - (\tilde{\beta} - \hat{\beta})' \hat{\Delta}^{-1} (\tilde{\beta} - \hat{\beta}), \tag{3.13}$$

$$\Psi_R = \frac{1}{T}\{\bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z]\}, \quad \hat{\sigma}_R^2 = y' \Lambda_R y, \quad \Lambda_R = \frac{1}{T}\bar{\mathrm{M}}[Z], \tag{3.14}$$

$\kappa_1 = (k_2 - G)/G,\ \kappa_2 = (T - k_1 - 2G)/G,\ \kappa_3 = \kappa_4 = T - k_1 - G$, and $\kappa_R = (T - k_1 - k_2 - G)/k_2$. Here, $\hat{u}$ is the vector of OLS residuals from equation (2.1) and $\hat{\sigma}^2$ is the corresponding OLS-based estimator of σ_u^2 (without correction for degrees of freedom), while $\tilde{u}$ is the vector of the 2SLS residuals and $\tilde{\sigma}^2$ the usual 2SLS-based estimator of σ_u^2; $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$, $\tilde{\sigma}_e^2$ and $\hat{\sigma}_R^2$ may be interpreted as alternative IV-based scaling factors. Note also that $P_1 P = P P_1 = P_1$, $M_1 M = M M_1 = M$, and

$$\begin{aligned} N_1 &= M_1 P = P M_1 = P M_1 P = M_1 P M_1 = N_1 M_1 = M_1 N_1 = N_1 N_1 \\ &= M_1 - M = P - P_1 = \bar{\mathrm{P}}[X] - \bar{\mathrm{P}}[X_1] = \bar{\mathrm{P}}[M_1 X_2]. \end{aligned} \tag{3.15}$$

Each one of the corresponding tests rejects H_0 when the statistic is "large". We also set

$$\hat{V} := MY, \quad \hat{\Sigma}_V := \frac{1}{T}\hat{V}'\hat{V}, \tag{3.16}$$

i.e. $\hat{\Sigma}_V$ is the usual sample covariance matrix of the LS residuals $(\hat{V})$ from the reduced-form linear model (2.7).

The tests differ through the use of different "covariance matrix" estimators. $\mathscr{H}_1$ uses two different estimators of σ_u^2, while the others resort to a single scaling factor (or estimator of σ_u^2). We think the expressions given here for $\mathscr{T}_l, l = 1, 2, 3, 4$, in (3.1) are easier to interpret than those of Wu (1973), and show more clearly the relation with Hausman-type tests. The statistic $\mathscr{H}_1$ can be interpreted as the statistic proposed by Hausman (1978), while $\mathscr{H}_2$ and $\mathscr{H}_3$ are sometimes interpreted as variants of $\mathscr{H}_1$ [see Staiger and Stock (1997) and Hahn et al. (2010)]. We use the above notations to better see the relation between Hausman-type tests and Wu-type tests. In particular, $\tilde{\Sigma}_3 = \hat{\Sigma}_2$ and $\tilde{\Sigma}_4 = \hat{\Sigma}_3$, so $\mathscr{T}_3 = (\kappa_3/T)\mathscr{H}_2$ and $\mathscr{T}_4 = (\kappa_4/T)\mathscr{H}_3$. Further, $\mathscr{T}_4$ is a nonlinear monotonic transformation of $\mathscr{T}_2$:

$$\mathscr{T}_4 = \frac{\kappa_4 \mathscr{T}_2}{\mathscr{T}_2 + \kappa_2} = \frac{\kappa_4}{(\kappa_2/\mathscr{T}_2) + 1}. \tag{3.17}$$

Despite these relations, the tests based on $\mathscr{T}_3$ and $\mathscr{H}_2$ are equivalent only if exact critical values are used, and similarly for the pairs $(\mathscr{T}_4, \mathscr{H}_3)$ and $(\mathscr{T}_2, \mathscr{T}_4)$. We are not aware of a simple equivalence between $\mathscr{H}_1$ and $\mathscr{T}_i, i = 1, 2, 3, 4$, and similarly between $\mathscr{T}_1$ and $\mathscr{H}_j, j = 1, 2, 3$.

The link between the formulation of Wu (1973) and the one above is discussed in Appendix A.[2] Condition (2.6) entails that $\hat{\Omega}_{IV}$, $\hat{\Omega}_{LS}$ and $\hat{\Sigma}_V$ are (almost surely) nonsingular, which in turn implies that $\hat{\Delta}$ is invertible; see Lemma A.1 in Appendix. In particular, it is of interest to observe that

$$\begin{aligned}\hat{\Delta}^{-1} &= \hat{\Omega}_{IV} + \hat{\Omega}_{IV}(\hat{\Omega}_{LS} - \hat{\Omega}_{IV})^{-1}\hat{\Omega}_{IV} = \hat{\Omega}_{IV} + \hat{\Omega}_{IV}\hat{\Sigma}_V^{-1}\hat{\Omega}_{IV} = \hat{\Omega}_{LS}\hat{\Sigma}_V^{-1}\hat{\Omega}_{LS} - \hat{\Omega}_{LS} \\ &= \frac{1}{T}Y'N_1\left[I_T + Y(Y'MY)^{-1}Y'\right]N_1Y = \frac{1}{T}Y'M_1[Y(Y'MY)^{-1}Y' - I_T]M_1Y\,, \end{aligned} \tag{3.18}$$

from which we see easily that $\hat{\Delta}^{-1}$ is positive definite. Further, $\hat{\Delta}^{-1}$ only depends on the least-squares residuals M_1Y and MY from the regressions of Y on X_1 and X respectively, and $\hat{\Delta}^{-1}$ can be bounded as follows:

$$\hat{\Omega}_{IV} \leq \hat{\Delta}^{-1} \leq \hat{\Omega}_{LS}\hat{\Sigma}_V^{-1}\hat{\Omega}_{LS} \tag{3.19}$$

so that

$$(\tilde{\beta} - \hat{\beta})'\hat{\Omega}_{IV}(\tilde{\beta} - \hat{\beta}) \leq (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) \leq (\tilde{\beta} - \hat{\beta})'\hat{\Omega}_{LS}\hat{\Sigma}_V^{-1}\hat{\Omega}_{LS}(\tilde{\beta} - \hat{\beta})\,. \tag{3.20}$$

To the best of our knowledge, the additive expressions in (3.18) are not available elsewhere.

Finite-sample distributional results are available for $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{R}$ when the disturbances u_t are i.i.d. Gaussian. If $u \sim N[0, \sigma^2 I_T]$ and X is independent of u, we have:

$$\mathscr{T}_1 \sim F(G, k_2 - G)\,, \quad \mathscr{T}_2 \sim F(G, T - k_1 - 2G)\,, \quad \mathscr{R} \sim F(k_2, T - k_1 - k_2 - G)\,, \tag{3.21}$$

under Assumptions 2.1 and 2.2 with $a = 0$. Since $\mathscr{T}_4$ and $\mathscr{H}_3$ are strictly monotonic transformations of $\mathscr{T}_2$, this also entails that the distribution functions and densities of $\mathscr{T}_4$ and $\mathscr{H}_3$ can be derived from the $F(G, T - k_1 - 2G)$ distribution. For large samples, we have under Assumptions 2.1 and 2.2 with $a = 0$ (along with standard asymptotic regularity conditions):

$$\mathscr{H}_i \xrightarrow{L} \chi^2(G)\,, i = 1, 2, 3 \text{ and } \mathscr{T}_l \xrightarrow{L} \chi^2(G)\,, l = 3, 4, \tag{3.22}$$

when $\text{rank}(\Pi_2) = G$., where $\xrightarrow{L}$ means convergence in distribution as $T \to \infty$.

Finite-sample distributional results are not available in the literature for $\mathscr{H}_i$, $i = 1, 2, 3$ and $\mathscr{T}_l$, $l = 3, 4$, even when errors are Gaussian and usual full identification assumptions are made. Of course, the same remark applies when usual conditions for identification fail [$\text{rank}(\Pi_2) < G$] or get close to do so – *e.g.*, some eigenvalues of $\Pi_2'\Pi_2$ are close to zero (weak identification) – and disturbances may not be Gaussian. This paper provides a formal characterization of the size and power of the tests when IVs may be arbitrary weak, with and without Gaussian errors.

[2] When the errors $U_1, \ldots, U_T$ are i.i.d. Gaussian [in which case Assumptions 2.4 and 2.5 hold], the $\mathscr{T}_2$ test of Wu (1973) can also be interpreted as the LM test of $a = 0$; see Smith (1983) and Engle (1982).

3.2. Regression-based formulations of exogeneity statistics

We now show that all the above test statistics can be computed from relatively simple linear regressions, which may be analytically revealing and computationally convenient. We consider again the regression of u on V in (2.10):

$$u = Va + e \tag{3.23}$$

for some constant vector $a \in \mathbb{R}^G$, where e has mean zero and variance σ_e^2, and is uncorrelated with V and X. We can write the structural equation (2.1) in three different ways as follows:

$$\begin{aligned} y &= Y\beta + X_1\gamma + \hat{V}a + e_* = \hat{Z}\theta + e_*\,, & (3.24)\\ y &= \hat{Y}\beta + X_1\gamma + \hat{V}b + e_* = Z_*\theta_* + e_*\,, & (3.25)\\ y &= Yb + X_1\bar{\gamma} + X_2\bar{a} + e = \bar{Z}_*\bar{\theta} + e\,, & (3.26) \end{aligned}$$

where

$$\hat{Z} = [Y, X_1, \hat{V}]\,,\ \theta = (\beta', \gamma', a')',\ Z_* = [\hat{Y}, X_1, \hat{V}]\,,\ \theta_* = (\beta', \gamma', b')',\ \bar{Z}_* = [Y, X_1, X_2]\,, \tag{3.27}$$

$$\bar{\theta} = (b', \bar{\gamma}', \bar{a}')',\ b = \beta + a,\ \bar{\gamma} = \gamma - \Pi_1 a, \quad \bar{a} = -\Pi_2\, a\,, \tag{3.28}$$

$$\hat{Y} = \bar{\mathrm{P}}[X]Y, \quad \hat{V} = \bar{\mathrm{M}}[X]Y\,, \quad e_* = \bar{\mathrm{P}}[X]Va + e\,. \tag{3.29}$$

Clearly, $\beta = b$ if and only if $a = 0$. Equations (3.23) - (3.26) show that the endogeneity of Y in (2.1) - (2.7) can be interpreted as an omitted-variable problem [for further discussion of this view, see Dufour (1979, 1987) and Doko Tchatoka and Dufour (2014)]. The inclusion of $\hat{V}$ in equations (3.24) - (3.25) may also be interpreted as an application of control function methods [see Wooldridge (2015)]. We also consider the intermediate regression:

$$y - Y\tilde{\beta} = X_1\bar{\gamma} + X_2\bar{a} + e_{**} = X\theta_{**} + e_{**} \tag{3.30}$$

where $\tilde{\beta}$ is the 2SLS estimator of β.

Let $\hat{\theta}$ be the OLS estimator of θ and $\hat{\theta}^0$ the restricted OLS estimator of θ under the constraint $H_0 : a = 0$ [in (3.24)], $\hat{\theta}_*$ the OLS estimator of θ_* and $\hat{\theta}_*^0$ the restricted OLS estimate of θ_* under $H_0^* : \beta = b$ [in (3.25)], $\check{\theta}$ the OLS estimate of $\bar{\theta}$ and $\check{\theta}^0$ the restricted OLS estimate of $\bar{\theta}$ under $\bar{H}_0 : \bar{a} = 0$ [in (3.26)]. Similarly, the OLS estimate of θ_{**} based on (3.30) is denoted $\hat{\theta}_{**}$, while $\hat{\theta}_{**}^0$ represents the corresponding restricted estimate under $\bar{H}_0 : \bar{a} = 0$. The sum of squared error functions associated with (3.24) - (3.26) are denoted:

$$S(\theta) = \|y - \hat{Z}\theta\|^2, \quad S_*(\theta_*) = \|y - Z_*\theta_*\|^2, \quad \bar{S}(\bar{\theta}) = \|y - \bar{Z}_*\bar{\theta}\|^2\,, \tag{3.31}$$

$$\tilde{S}(\theta_{**}) = \|y - Y\tilde{\beta} - X\theta_{**}\|^2\,. \tag{3.32}$$

Using $Y = \hat{Y} + \hat{V}$, we see that:

$$S(\hat{\theta}) = S_*(\hat{\theta}_*) = \bar{S}(\check{\theta}^0)\,, \quad S(\hat{\theta}^0) = S_*(\hat{\theta}_*^0) = \tilde{S}(\hat{\theta}_{**}^0)\,, \tag{3.33}$$

$$S(\hat{\theta}) = T\,\tilde{\sigma}_2^2\,, \quad S(\hat{\theta}^0) = T\,\hat{\sigma}^2\,, \quad S_*(\hat{\theta}_*^0) = T\,\tilde{\sigma}^2\,, \quad \tilde{S}(\hat{\theta}_{**}) = T\,\tilde{\sigma}_e^2\,. \tag{3.34}$$

We then get the following expressions for the statistics in (3.1) - (3.3) :

$$\mathscr{T}_1 = \kappa_1 \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0) - \tilde{S}(\hat{\theta}_{**})} \right) = \kappa_1 \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{\tilde{S}(\hat{\theta}_{**}^0) - \tilde{S}(\hat{\theta}_{**})} \right), \tag{3.35}$$

$$\mathscr{T}_2 = \kappa_2 \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta})} \right), \quad \mathscr{T}_3 = \kappa_3 \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)} \right), \quad \mathscr{T}_4 = \kappa_4 \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta}^0)} \right), \tag{3.36}$$

$$\mathscr{H}_2 = T \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)} \right), \quad \mathscr{H}_3 = T \left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta}^0)} \right), \tag{3.37}$$

$$\mathscr{R} = \kappa_R\, [\bar{S}(\check{\theta}^0) - \bar{S}(\check{\theta})]/\bar{S}(\check{\theta})\,. \tag{3.38}$$

Details on the derivation of the above formulas are given in Appendix B.

(3.37) - (3.38) provide simple regression formulations of the DWH and RH statistics in terms of restricted and unrestricted sum of squared errors in linear regressions. However, we did not find such a simple expression for the Hausman statistic $\mathscr{H}_1$. While DWH-type tests consider the null hypothesis $H_0 : a = 0$, the RH test focuses on the null hypothesis $H_0^* : \bar{a} = -\Pi_2\, a = 0$. If $\text{rank}(\Pi_2) = G$, we have : $a = 0$ if and only if $\bar{a} = 0$. However, if $\text{rank}(\Pi_2) < G$, $\bar{a} = 0$ does not imply $a = 0$: H_0 entails H_0^*, but the converse does not hold in this case.

The regression interpretation of the $\mathscr{T}_2$ and $\mathscr{H}_3$ statistics was mentioned earlier in Dufour (1979, 1987) and Nakamura and Nakamura (1981). The $\mathscr{R}$ statistic was also derived as a standard regression test by Revankar and Hartley (1973). To our knowledge, the other regression interpretations given here are not available elsewhere.

4. Incomplete models and pivotal properties

In this section, we study the finite-sample null distributions of DWH-type and RH exogeneity tests under Assumption 2.1, allowing for the possibility of identification failure (or weak identification) and model incompleteness. The proofs of these results rely on two lemmas of independent interest (Lemmas C.1 - C.2) given in Appendix.

4.1. Distributions of test statistics under exogeneity

We first show that the exogeneity test statistics in (3.1) - (3.3) can be rewritten as follows, whether the null hypothesis holds or not.

Proposition 4.1 QUADRATIC-FORM REPRESENTATIONS OF EXOGENEITY STATISTICS. *The* *exogeneity test statistics in* (3.1) - (3.3) *can be expressed as follows :*

$$\mathscr{T}_l = \kappa_l \left(\frac{y'\Psi_0\, y}{y'\Lambda_l\, y}\right),\ \textit{for}\ l = 1, 2, 3, 4, \tag{4.1}$$

$$\mathscr{H}_1 = T\,(y'\Psi_1[y]\,y) = T\,(C_1 y)'\left[(y'\Lambda_3\, y)\,\hat{\Omega}_{IV}^{-1} - (y'\Lambda_4\, y)\,\hat{\Omega}_{LS}^{-1}\right]^{-1}(C_1 y)\,, \tag{4.2}$$

$$\mathscr{H}_2 = T\left(\frac{y'\Psi_0\, y}{y'\Lambda_3\, y}\right),\quad \mathscr{H}_3 = T\left(\frac{y'\Psi_0\, y}{y'\Lambda_4\, y}\right),\quad \mathscr{R} = \kappa_R\left(\frac{y'\Psi_R\, y}{y'\Lambda_R\, y}\right), \tag{4.3}$$

where the notation in (3.4) - (3.14) *is used,*

$$\Lambda_1 = \frac{1}{T}N_1\,\bar{\mathrm{M}}[N_1 Y]\,N_1\,,\quad \Lambda_2 = M_1\left(\frac{1}{T}\bar{\mathrm{M}}[M_1 Y] - \Psi_0\right)M_1\,, \tag{4.4}$$

$$\Lambda_3 = \frac{1}{T}M_1\,N_2' N_2 M_1\,,\quad \Lambda_4 = \frac{1}{T}\bar{\mathrm{M}}[\bar{Y}] = \frac{1}{T}M_1\bar{\mathrm{M}}[M_1 Y]M_1\,, \tag{4.5}$$

$$\Psi_1[y] = C_1'\hat{\Sigma}_1^{-1}C_1 = C_1'\left[(y'\Lambda_3\, y)\hat{\Omega}_{IV}^{-1} - (y'\Lambda_4\, y)\hat{\Omega}_{LS}^{-1}\right]^{-1}C_1\,, \tag{4.6}$$

and Ψ_0*,* N_2*,* C_1 *are defined as in Lemma C.1.*

The following theorem characterizes the distributions of all exogeneity statistics under the null hypothesis of exogeneity ($H_0 : a = 0$).

Theorem 4.2 NULL DISTRIBUTIONS OF EXOGENEITY STATISTICS. *Under the model described by* (2.1) - (2.6)*, suppose Assumption 2.1 holds. If* $H_0 : a = 0$ *also holds, then the test statistics defined in* (3.1) - (3.3) *have the following representations :*

$$\mathscr{T}_l = \kappa_l \left(\frac{\varepsilon'\Psi_0\,\varepsilon}{\varepsilon'\Lambda_l\,\varepsilon}\right),\ \textit{for}\ l = 1, 2, 3, 4, \tag{4.7}$$

$$\mathscr{H}_1 = T\,(\varepsilon'\Psi_1[\varepsilon]\,\varepsilon) = T\,(C_1\varepsilon)'\left[(\varepsilon'\Lambda_3\,\varepsilon)\,\hat{\Omega}_{IV}^{-1} - (\varepsilon'\Lambda_4\,\varepsilon)\,\hat{\Omega}_{LS}^{-1}\right]^{-1}(C_1\varepsilon)\,, \tag{4.8}$$

$$\mathscr{H}_2 = T\left(\frac{\varepsilon'\Psi_0\,\varepsilon}{\varepsilon'\Lambda_3\,\varepsilon}\right),\quad \mathscr{H}_3 = T\left(\frac{\varepsilon'\Psi_0\,\varepsilon}{\varepsilon'\Lambda_4\,\varepsilon}\right),\quad \mathscr{R} = \kappa_R\left(\frac{\varepsilon'\Psi_R\,\varepsilon}{\varepsilon'\Lambda_R\,\varepsilon}\right), \tag{4.9}$$

where Ψ_0*,* $\Lambda_1, \ldots, \Lambda_4$*,* Ψ_1*,* Ψ_R *and* Λ_R *are defined as in Proposition 4.1. If Assumption 2.2 also holds, the distributions of the test statistics* $\mathscr{T}_1$*,* $\mathscr{T}_2$*,* $\mathscr{T}_3$*,* $\mathscr{T}_4$*,* $\mathscr{H}_1$*,* $\mathscr{H}_2$*,* $\mathscr{H}_3$ *and* $\mathscr{R}$*, conditional on* $\bar{X}$ *and* Y*, only depend on the conditional distribution of* ε *given* $\bar{X}$*, as specified in Assumption 2.1, and the values of* Y *and* X*.*

The last statement of Theorem 4.2 comes from the fact that the weighting matrices defined in (4.4) - (4.6) only depend on X, Y and ε. Given X and Y, the null distributions of the exogeneity test statistics only depend on the distribution of ε : provided the distribution of $\varepsilon\,|\,\bar{X}$ can be simulated, exact tests can be obtained through the MCT method [see Section 4.2]. Furthermore, the tests obtained in this way are robust to weak instruments in the sense that the level is controlled even if identification fails (or is weak). This result holds even if the distribution of $\varepsilon\,|\,\bar{X}$ does not have moments (the Cauchy distribution, for example). This may be useful, for example, in financial models with fat-tailed error distributions, such as the Student t distribution. There is no further restriction on the distribution of $\varepsilon\,|\bar{X}$. For example, the distribution of $\varepsilon\,|\bar{X}$ may depend on $\bar{X}$, provided it can be simulated. Indeed, Assumption 2.2 could be replaced by a different assumption on the conditional distribution of ε given Y and $\bar{X}$. It is important to note that Theorem 4.2 does not *stricto sensu* involve a conditioning argument. Instead, it provides a characterization of the form of the test statistics as functions of ε, Y and X when $a = 0$. Once we condition on Y and $\bar{X}$, *conditional* distributions can be simulated, yielding tests with the desired level conditionally (hence also unconditionally).

It is interesting to observe that the distribution of V plays no role here, so the vectors $V_1, \ldots, V_T$ may follow arbitrary distributions with unspecified heterogeneity (or heteroskedasticity) and serial dependence. In addition to finite-sample validity of all the exogeneity tests in the presence of identification failure (or weak identification), Theorem 4.2 entails robustness to *incomplete reduced forms* and *instrument exclusion* under the null hypothesis of exogeneity. No further information is needed on the form of the reduced form for Y in (2.2) : $g(\cdot)$ can be an unspecified nonlinear function, $\Pi = [\Pi_1\,,\Pi_2\,]$ an unknown parameter matrix, and V may follow an arbitrary distribution. This result extends to the exogeneity tests the one given in Dufour and Taamouti (2007) on Anderson-Rubin-type tests (for structural coefficients).

As long as the distribution of ε (given $\bar{X}$ and Y) can be simulated, all tests remain valid under H_0, and test sizes are controlled conditional on $\bar{X}$ and Y, hence also unconditionally. In particular, Monte-Carlo test procedures remain valid even if the instrument matrix X_3 is not used by the test statistics. A similar property is underscored in Dufour and Taamouti (2007) for Anderson-Rubin tests in linear structural equation models. This observation is also useful to allow for models with structural breaks in the reduced form : exogeneity tests remain valid in such contexts without knowledge of the form and timing of breaks. In such contexts, inference on the structural form may be more reliable than inference on the reduced form, a question of great relevance for macroeconomic models; see Dufour et al. (2013). However, although the exclusion of instruments does not affect the null distributions of exogeneity test statistics, it may lead to power losses when the missing information is important.

4.2. Exact Monte Carlo exogeneity tests

To implement the exact MC exogeneity tests of H_0 with level α $(0 < \alpha < 1)$, we suggest the following methodology; for a more general discussion, see Dufour (2006). Suppose that the conditional distribution of ε (given $\bar{X}$) is continuous, so that the conditional distribution, given $\bar{X}$, of all exogeneity statistics is also continuous. Let $\mathscr{W}$ denotes any of the DWH and RH statistic in (3.1) - (3.3). We can then proceed as follows :

1. choose α^* and N so that
$$\alpha = \frac{I[\alpha^* N] + 1}{N+1} \tag{4.10}$$
where for any nonnegative real number x, $I[x]$ is the largest integer less than or equal to x;

2. compute the test statistic $\mathscr{W}^{(0)}$ based on the observed data;

3. generate N *i.i.d.* error vectors $\varepsilon^{(j)} = [\varepsilon_1^{(j)}, \ldots, \varepsilon_T^{(j)}]'$, $j = 1, \ldots, N$, according to the specified distribution of $\varepsilon \,|\bar{X}$, and compute the corresponding statistics $\mathscr{W}^{(j)}$, $j = 1, \ldots, N$, following Theorem 4.2; the distribution of each statistic does not depend on β_0 under the null hypothesis;

4. compute the empirical distribution function based on $\mathscr{W}^{(j)}$, $j = 1, \ldots, N$,

$$\hat{F}_N(x) = \frac{\sum_{j=1}^{N} \mathbb{1}[\mathscr{W}^{(j)} \leq x]}{N+1} \tag{4.11}$$

   or, equivalently, the simulated p-value function

$$\hat{p}_N[x] = \frac{1 + \sum_{j=1}^{N} \mathbb{1}[\mathscr{W}^{(j)} \geq x]}{N+1} \tag{4.12}$$

   where $\mathbb{1}[C] = 1$ if condition C holds, and $\mathbb{1}[C] = 0$ otherwise;

5. reject the null hypothesis of exogeneity, H_0, at level α when $\mathscr{W}^{(0)} \geq \hat{F}_N^{-1}(1-\alpha^*)$, where $\hat{F}_N^{-1}(q) = \inf\{x : \hat{F}_N(x) \geq q\}$ is the generalized inverse of $\hat{F}_N(\cdot)$, or (equivalently) when $\hat{p}_N[\mathscr{W}^{(0)}] \leq \alpha$.

Under H_0,

$$\mathbb{P}\big[\hat{p}_N[\mathscr{W}^{(0)}] \leq \alpha\big] = \alpha \tag{4.13}$$

so that we have a test with level α. The property given by (4.13) is a finite-sample validity result which holds irrespective of the sample size T, and no asymptotic assumption is required. If the distributions of the statistics are not continuous, the MCT procedure can easily be adapted by using "tie-breaking" method described in Dufour (2006).[3]

It is important to note here that the distributions of the exogeneity test statistics in Theorem 4.2 generally depend on the specific "instrument matrix" X used by the tests (especially when ε is not Gaussian), so no general valid "critical value" (independent of X) is available. The MCT procedure transparently controls the level of the test irrespective of this complication, so there is *no need to compute critical values*.

The finite-sample validity of the MC exogeneity tests described here is allowed by conditioning on $\bar{X}$ and Y under the null hypothesis. This form of conditioning allows one to avoid specifying a marginal model for $\bar{X}$ and Y, in the same way that conditioning on a regressor matrix in linear regression provides valid tests without the need to specify a model for the regressors. So the procedures so obtained are "nonparametric" with respect to the distribution of Y. This type of conditioning is quite distinct from conditioning on sufficient statistics [as suggested, for example, by Moreira (2003)]. Since the distribution of Y is not specified, (complete) sufficient statistics for the parameters of the Y distribution are not available here, so conditional likelihood ratio tests of this type are not feasible in our setup. If such a model for Y were introduced, more powerful test procedures might be achieved,

[3]Without correction for continuity, the algorithm proposed for statistics with continuous distributions yields a conservative test, *i.e.* the probability of rejection under the null hypothesis is not larger than the nominal level (α). Further discussion of this feature is available in Dufour (2006).

but their validity and power properties would depend crucially on the additional assumptions. Of course, in specific contexts, this may provide an interesting modification of DWH tests. We leave this possibility to future research.

4.3. Asymptotic validity of the Monte Carlo exogeneity tests

Even though the main focus of this paper consists in presenting a finite-sample distributional theory for exogeneity test statistics, it will be useful to briefly discuss the validity of the proposed procedures under limiting assumptions. Such assumptions are weaker in some ways (for example, through the imposition of less specific distributional assumptions), and stronger in other ways (thorough assumptions on the existence of moments and other restrictions needed for central limit theorems). To do this, we consider the following additional high-level regularity assumptions on the model variables and errors, where "$\xrightarrow{p}$" stands for convergence in probability, "$\xrightarrow{d}$" means convergence in distribution, $vec(\cdot)$ is the column vectorization operator, and $\otimes$ symbolizes the Kronecker (tensor) product.

Assumption 4.1 CONVERGENCE ASSUMPTIONS. *When the sample size T converges to infinity, the following convergence properties hold jointly :*

$$\frac{1}{T}\bar{X}'[u, V] \xrightarrow{p} 0\,,\ \frac{1}{T}[u, V]'[u, V] \xrightarrow{p} \Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_V \end{bmatrix} > 0\,,\ \frac{1}{T}\bar{X}'\bar{X} \xrightarrow{p} \Sigma_{\bar{X}} > 0\,, \tag{4.14}$$

$$\frac{1}{\sqrt{T}}\mathrm{vec}[\bar{X}'(u, V)] \xrightarrow{d} \mathrm{vec}(\psi_{\bar{X}u},\, \psi_{\bar{X}V}) \sim N[0, \Sigma \otimes \Sigma_{\bar{X}}]\,, \tag{4.15}$$

$$\sqrt{T}\left(\frac{1}{T}V'u - \sigma_{Vu}\right) \xrightarrow{d} \psi_{Vu} \sim N[0, \Omega_{Vu}]\,, \tag{4.16}$$

where $\Omega_{Vu} = \sigma_u^2\Sigma_V$ when $\sigma_{Vu} = 0$.

Under the model described by (2.1) - (2.6) with Assumptions 2.4 and 4.1, we have :

$$\mathscr{T}_1 \xrightarrow{d} F(G, k_2 - G)\,,\ \mathscr{T}_2 \xrightarrow{d} \frac{1}{G}\chi^2(G)\,,\ \mathscr{T}_4 \xrightarrow{d} \chi^2(G)\,,\ \mathscr{H}_3 \xrightarrow{d} \chi^2(G)\,,\ \mathscr{R} \xrightarrow{d} \frac{1}{k_2}\chi^2(k_2), \tag{4.17}$$

when $\sigma_{Vu} = 0$, irrespective of whether identification is strong or weak; see Staiger and Stock (1997) [for $\mathscr{H}_3$] and Doko Tchatoka and Dufour (2016) [for all statistics].[4] Following Dufour (2006), it is straightforward to see that the MCT procedure described in Section 4.2 remain asymptotically valid for the statistics $\mathscr{H}_3$, $\mathscr{R}$, $\mathscr{T}_1$, $\mathscr{T}_2$, and $\mathscr{T}_4$. However, the null limiting asymptotic distributions of $\mathscr{H}_i, i = 1, 2$, and $\mathscr{T}_3$ are nonstandard under weak IVs [see Doko Tchatoka and Dufour (2016)], but they can be simulated under Assumption 4.1. Getting an asymptotic theory requires

[4] It is worth noting that the results of Staiger and Stock (1997) are derived under the assumption that the reduced-form parameter matrix Π in (2.7) is local-to-zero. The results presented in Doko Tchatoka and Dufour (2016) also allow for cases where Π is fixed and has a reduced rank.

additional (or different assumptions), such as Assumption 4.1, but as long as they include the finite-sample assumptions used by the MCT procedure [including possibly assumptions on moments, like $\mathbb{E}(\varepsilon|\bar{X}) = 0$], the MCT procedures of Section 4.2 also remain asymptotically valid for these statistics.

An important implication of the above asymptotic validity is that the MCT procedure yields tests with correct size asymptotically, even when the distribution of $\varepsilon \mid \bar{X}$ was incorrectly specified during the construction of MCT p-values, provided that the high-level Assumption 4.1 holds. The simulation experiments (Section 7) show that the MC tests have an overall good performance under the misspecification of the conditional distribution of ε given $\bar{X}$, even with a sample size $T = 50$, with or without weak instruments (see Table 4).[5]

5. Block-triangular invariance and exogeneity canonical form

In this section, we establish invariance results for exogeneity tests which will be useful to study the distributions of the test statistics under the alternative hypothesis. This basic invariance property is given by the following proposition.

Proposition 5.1 BLOCK-TRIANGULAR INVARIANCE OF EXOGENEITY TESTS. *Let*

$$R = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \tag{5.1}$$

be a lower block-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* (3.1) - (3.14)*, the statistics* $\mathscr{T}_i$ $(i = 1, 2, 3, 4)$*,* $\mathscr{H}_j$ $(j = 1, 2, 3)$ *and* $\mathscr{R}$ *do not change.*

The above result is purely algebraic, so no statistical assumption is needed. However, when it is combined with our statistical model, it has remarkable consequences on the properties of exogeneity tests. For example, if the reduced-form errors $V_1, \ldots, V_T$ for Y have the same nonsingular covariance matrix Σ, the latter can be eliminated from the distribution of the test statistic by choosing R_{22} so that $R_{22}'\Sigma R_{22} = I_G$. This entails that the distributions of the exogeneity statistics do not depend on Σ under both the null and the alternative hypotheses.

Consider now the following transformation matrix :

$$R = \begin{bmatrix} 1 & 0 \\ -(\beta + a) & I_G \end{bmatrix}. \tag{5.2}$$

[5] For cases where Gaussian-based MC tests are not reliable even asymptotically (e.g., due to heavy tails), a more thorough modelling of the error distribution (at least, under the null hypothesis) may be needed, including goodness-of-fit tests to identify reasonable error distributions along with procedures for taking into account the additional parameter uncertainty introduced by this process. Interestingly, Theorem 4.2 can still play a central role in implementing exogeneity tests with non-Gaussian errors. For examples of such procedures (in different models), see Dufour, Farhat, Gardiol and Khalaf (1998), Dufour, Khalaf and Beaulieu (2003), Beaulieu, Dufour and Khalaf (2005, 2007, 2009, 2013). Setting up an integrated procedure for doing this in the present context would go beyond the scope of this paper.

Then, we have $[y^{*}, Y^{*}] = [y, Y]R$ with

$$
\begin{aligned}
y^{*} &= y - Y(\beta + a) = Y\beta + X_1\gamma + Va + e - Y(\beta + a) = \mu_{y*}(a) + e\,, && (5.3)\\
Y^{*} &= Y && (5.4)
\end{aligned}
$$

where $\mu_{y*}(a)$ is a $T \times 1$ vector such that

$$\mu_{y^*}(a) = X_1\gamma + [V - g(X_1, X_2, X_3, V, \bar{\Pi})]a\,. \tag{5.5}$$

The (invertible) transformation (5.3) - (5.4) yields the following "latent reduced-form" representation :

$$y^{*} = X_1\gamma + [V - g(X_1, X_2, X_3, V, \bar{\Pi})]a + e\,, \tag{5.6}$$

$$Y = g(X_1, X_2, X_3, V, \bar{\Pi})\,. \tag{5.7}$$

We say "latent" because the function $g(\cdot)$ and the variables X_3 are unknown or unspecified. An important feature here is that the endogeneity parameter a can be isolated from other model parameters. This will allow us to get relatively simple characterizations of the power of exogeneity tests. For this reason, we call (5.6) - (5.7), the "exogeneity canonical form" associated with model (2.1) - (2.2) along with Assumption 2.1.

In the important case where reduced-form error linear separability holds (Assumption 2.7) in addition to (2.1) - (2.2), we can write

$$Y = g(X_1, X_2, X_3, \bar{\Pi}) + V = \mu_Y + V \tag{5.8}$$

which, by (2.1), entails

$$y = \mu_y(a) + (u + V\beta) = \mu_y(a) + v \tag{5.9}$$

where μ_Y is a $T \times G$ matrix and μ_y is a $T \times 1$ vector, such that

$$\mu_Y = g(X_1, X_2, X_3, \bar{\Pi})\,, \quad \mu_y(a) = g(X_1, X_2, X_3, \bar{\Pi})\beta + X_1\gamma, \tag{5.10}$$

$$v = u + V\beta = e + V(\beta + a)\,. \tag{5.11}$$

Then

$$\mu_{y^*}(a) = \mu_y(a) - \mu_Y(\beta + a) = X_1\gamma - g(X_1, X_2, X_3, \bar{\Pi})a \tag{5.12}$$

does not depend on V, and the exogeneity canonical form is :

$$y^{*} = X_1\gamma - g(X_1, X_2, X_3, \bar{\Pi})a + e\,, \tag{5.13}$$

$$Y = g(X_1, X_2, X_3, \bar{\Pi}) + V\,. \tag{5.14}$$

6. Power

In this section, we provide characterizations of the power of exogeneity tests. We first consider the general case where only Assumption 2.1 is added to the basic setup (2.1) - (2.6). To simplify the exposition, we use the following notation : for any $T \times 1$ vector x and $T \times T$ matrix A, we set

$$S_T[x, A] = T\, x' A\, x\,. \tag{6.1}$$

Theorem 6.1 EXOGENEITY TEST DISTRIBUTIONS UNDER THE ALTERNATIVE HYPOTHESIS. *Under the model described by* (2.1) - (2.6)*, suppose Assumption 2.1 holds. Then the test statistics defined in* (3.1) - (3.3) *have the following representations :*

$$\mathscr{T}_l = \kappa_l \left(\frac{S_T[u(\bar{a}), \Psi_0]}{S_T[u(\bar{a}), \Lambda_l]} \right), \quad \textit{for } l = 1, 2, 3, 4, \tag{6.2}$$

$$\mathscr{H}_1 = T\,\{u(\bar{a}\,)' \Psi_1[u(\bar{a}\,)]\, u(\bar{a}\,)\}, \quad \mathscr{H}_2 = T\, \left(\frac{S_T[u(\bar{a}\,), \Psi_0]}{S_T[u(\bar{a}\,), \Lambda_3]} \right), \quad \mathscr{H}_3 = T\, \left(\frac{S_T[u(\bar{a}\,), \Psi_0]}{S_T[u(\bar{a}\,), \Lambda_4]} \right), \tag{6.3}$$

$$\mathscr{R} = \kappa_R \left(\frac{S_T[u(\bar{a}\,), \Psi_R]}{S_T[u(\bar{a}\,), \Lambda_R]} \right), \tag{6.4}$$

where $u(\bar{a}\,) = V\bar{a} + \varepsilon$, $\bar{a}\; = \sigma(\bar{X})^{-1}a$,

$$\Psi_1[u(\bar{a}\,)] = C_1' \big(S_T[u(\bar{a}\,), \Lambda_3]\, \hat{\Omega}_{IV}^{-1} - S_T[u(\bar{a}\,), \Lambda_4]\, \hat{\Omega}_{LS}^{-1} \big)^{-1} C_1 \tag{6.5}$$

and C_1, Ψ_0, Ψ_1, Ψ_R, Λ_R, $\Lambda_1, \ldots, \Lambda_4$ *are defined as in Theorem 4.2. If Assumption 2.6 also holds, the distributions of the test statistics* (*conditional on* $\bar{X}$) *depend on a only through* $\bar{a}$ *in* $u(\bar{a}\,)$.

By Theorem 6.1, the distributions of all the exogeneity statistics depend on a, though possibly in a rather complex way (especially when disturbances follow non-Gaussian distributions). If the distribution of ε does not depend on $\bar{a}$ – as would be typically the case – power depends on the way the distributions of the quadratic forms $S_T[u(\bar{a}\,), \Psi_i]$ and $S_T[u(\bar{a}\,), \Lambda_j]$ in (6.2) - (6.4) are modified when the value of $\bar{a}$ changes. Both the numerator and the denominator of the statistics in Theorem 6.1 may follow different distributions, in contrast to what happens in standard F tests in the classical linear model.

The power characterization given by Theorem 6.1 does not provide a clear picture of the parameters which determine the power of exogeneity tests. This can be done by exploiting the invariance result of Proposition 5.1, as follows.

Theorem 6.2 INVARIANCE-BASED DISTRIBUTIONS OF EXOGENEITY STATISTICS. *Under the model described by* (2.1) - (2.6)*, suppose Assumption 2.1 holds. Then the test statistics defined in* (3.1) - (3.3) *have the following representations :*

$$\mathscr{T}_l = \kappa_l \left(\frac{S_T[y_*^{\perp}(\bar{a}\,), \Psi_0]}{S_T[y_*^{\perp}(\bar{a}\,), \Lambda_l]} \right), \quad \textit{for } l = 1, 2, 3, 4, \tag{6.6}$$

$$\mathscr{H}_1 = S_T\left[y_*^{\perp}(\bar{a}\,),\, \Psi_1[y_*^{\perp}(\bar{a}\,)]\right], \quad \mathscr{H}_2 = T\left(\frac{S_T[y_*^{\perp}(\bar{a}\,),\, \Psi_0]}{S_T[y_*^{\perp}(\bar{a}\,),\, \Lambda_3]}\right), \tag{6.7}$$

$$\mathscr{H}_3 = T\left(\frac{S_T[y_*^{\perp}(\bar{a}\,),\, \Psi_0]}{S_T[y_*^{\perp}(\bar{a}\,),\, \Lambda_4]}\right), \quad \mathscr{R} = \kappa_R\left(\frac{S_T[y_*^{\perp}(\bar{a}\,),\, \Psi_R]}{S_T[y_*^{\perp}(\bar{a}\,),\, \Lambda_R]}\right), \tag{6.8}$$

where

$$y_*^{\perp}(\bar{a}\,) = \bar{\mu}_{y*}^{\perp}(\bar{a}\,) + M_1\varepsilon, \tag{6.9}$$

$$\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}, \quad \bar{a} = \sigma(\bar{X})^{-1}a\,, \tag{6.10}$$

$$\Psi_1[y_*^{\perp}(\bar{a}\,)] = C_1'\left(S_T[y_*^{\perp}(\bar{a}\,),\, \Lambda_3]\,\hat{\Omega}_{IV}^{-1} - S_T[y_*^{\perp}(\bar{a}\,),\, \Lambda_4]\,\hat{\Omega}_{LS}^{-1}\right)^{-1}C_1\,, \tag{6.11}$$

and C_1, Ψ_0, Ψ_1, Ψ_R, Λ_R, $\Lambda_1, \ldots, \Lambda_4$ are defined as in Theorem 4.2. *If Assumption 2.6 also holds, the distributions of the test statistics (conditional on $\bar{X}$ and V) depend on a only through $\bar{\mu}_{y*}^{\perp}(\bar{a}\,)$ in $y_*^{\perp}(\bar{a}\,)$. If Assumption 2.7 also holds,*

$$\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = -M_1\, g(X_1, X_2, X_3, \bar{\Pi})\,\bar{a}\,. \tag{6.12}$$

Following Theorem 6.2, the powers of the different exogeneity tests are controlled by $\bar{\mu}_{y*}^{\perp}(\bar{a}\,)$ in (6.10). Clearly $a = 0$ entails $\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = 0$, which corresponds to the distribution under the null hypothesis [under Assumption 2.6]. Note however, the latter property also holds when

$$M_1\,[V - g(X_1, X_2, X_3, V, \bar{\Pi})] = 0 \tag{6.13}$$

even if $a \neq 0$, meaning that all DWH and RH tests may have trivial power even under a large deviation from the null (*i.e.*, $a \neq 0$ is large).

Under Assumption 2.7, V is evacuated from $\bar{\mu}_{y*}^{\perp}(\bar{a}\,)$ as given by (6.12). If Assumptions 2.6 and 2.7 hold, power is determined by this parameter. $\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = 0$ (trivial power) when $a = 0$, but also when X_1 and $g(X_1, X_2, X_3, \bar{\Pi})$ are orthogonal. Note also the norm of $\bar{\mu}_{y*}^{\perp}(\bar{a}\,)$ shrinks when $\sigma(\bar{X})$ increases, so power decreases when the variance of ε_t increases (as expected). Under Assumption 2.7, conditioning on $\bar{X}$ and V also becomes equivalent to conditioning on $\bar{X}$ and Y.

Consider the special case of a complete linear model where equations (2.7) and (2.8) hold. We then have :

$$g(X_1, X_2, X_3, \bar{\Pi}) = X_1\Pi_1 + X_2\Pi_2\,, \quad \mu_{y*}^{\perp}(\bar{a}) = -M_1X_2\Pi_2\,\bar{a}\,. \tag{6.14}$$

When $\Pi_2 = 0$ (complete non-identification of model parameters) or $M_1X_2 = 0$ (X_2 perfectly collinear with X_1), or more generally when $M_1X_2\Pi_2 = 0$, we have $\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = 0$. Then, under Assumption 2.6, the distributions of the exogeneity test statistics do not depend on a, and the power function is flat (with respect to a). This explains why in the simulation exercise (see Tables 1-4), the rejection frequencies of all tests do not significantly exceed the nominal 5% level for all values of λ when $\eta_1 \in \{0, 0.01\}$, since these correspond to cases where $\bar{\mu}_{y*}^{\perp}(\bar{a}\,) = 0$ is zero (or close to) zero.

Of course, all DWH and RH tests exhibit low power under local violation of endogeneity (*i.e.*, small values of a). As seen from the linear reduced-form in (6.14), a small deviation from the null hypothesis ($\bar{a}$ close to zero) entails a small value of $\bar{\mu}_{y*}^{\perp}(\bar{a})$, which leads to low power for all exogeneity tests. However, if $M_1 X_2 \Pi_2 \neq 0$, $\bar{\mu}_{y*}^{\perp}(\bar{a})$ can take arbitrarily large values for appropriate values of $\bar{a}$, so that exogeneity tests can have power against endogeneity.

Theorem 6.2 provides a conditional power characterization [given $\bar{X}$ and V (or Y)]. Even though the level of the test does not depend on the distribution of V, power typically depends on the distribution of V. Unconditional power functions can be obtained by averaging over V, but this requires formulating specific assumptions on the distribution of V.

When the disturbances $\varepsilon_1, \ldots, \varepsilon_T$ are i.i.d. Gaussian, it is possible to express the power function in terms of non-central chi-square distributions. We denote by $\chi^2[n;\delta]$ the non-central chi-square distribution with n degrees of freedom and noncentrality parameter δ, and by $F[n_1, n_2;\, \delta_1, \delta_2]$ the doubly noncentral F-distribution with degrees of freedom (n_1, n_2) and noncentrality parameters (δ_1, δ_2), *i.e.* $F \sim F[n_1, n_2;\, \delta_1, \delta_2]$ means that F can be written as $F = [Q_1/n_1]\,/\,[Q_2/n_2]$ where Q_1 and Q_2 are two independent random variables such that $Q_1 \sim \chi^2[n_1;\delta_1]$ and $Q_2 \sim \chi^2[n_2;\delta_2]$; see Johnson, Kotz and Balakrishnan (1995, Ch. 30). When $\delta_2 = 0$, $F \sim F[n_1, n_2;\, \delta_1]$ the usual noncentral F-distribution.

Theorem 6.3 INVARIANCE-BASED DISTRIBUTIONS OF EXOGENEITY STATISTICS COMPONENTS WITH GAUSSIAN ERRORS. *Under the model described by* (2.1) - (2.6)*, suppose Assumptions 2.1 and 2.2 hold. If $\varepsilon \sim \mathrm{N}[0, I_T]$, then, conditional on $\bar{X}$ and V, we have :*

$$S_T[y_*^{\perp}(\bar{a}\,), \Psi_0] \sim \chi^2[G;\, \delta(\bar{a}, \Psi_0)]\,, \quad S_T[y_*^{\perp}(\bar{a}\,), \Lambda_1] \sim \chi^2[k_2 - G;\, \delta(\bar{a}, \Lambda_1)]\,, \tag{6.15}$$

$$S_T[y_*^{\perp}(\bar{a}\,), \Lambda_2] \sim \chi^2[T - k_1 - 2G;\, \delta(\bar{a}, \Lambda_2)]\,, \quad S_T[y_*^{\perp}(\bar{a}\,), \Lambda_4] \sim \chi^2[T - k_1 - G;\, \delta(\bar{a}, \Lambda_4)]\,, \tag{6.16}$$

$$S_T[y_*^{\perp}(\bar{a}\,), \Psi_R] \sim \chi^2[k_2;\, \delta(\bar{a}, \Psi_R)]\,, \quad S_T[y_*^{\perp}(\bar{a}\,), \Lambda_R] \sim \chi^2[T - k_1 - k_2 - G;\, \delta(\bar{a}, \Lambda_R)]\,, \tag{6.17}$$

where

$$\delta(\bar{a}, \Psi_0) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Psi_0]\,, \quad \delta(\bar{a}, \Lambda_1) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Lambda_1]\,, \tag{6.18}$$

$$\delta(\bar{a}, \Lambda_2) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Lambda_2]\,, \quad \delta(\bar{a}, \Lambda_4) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Lambda_4]\,, \tag{6.19}$$

$$\delta(\bar{a}, \Psi_R) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Psi_R]\,, \quad \delta(\bar{a}, \Lambda_R) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}\,), \Lambda_R]\,, \tag{6.20}$$

and the other symbols are defined as in Theorem 6.2. Further, conditional on $\bar{X}$ and V, the random variable $S_T[y_*^{\perp}(\bar{a}\,), \Psi_0]$ is independent of $S_T[y_*^{\perp}(\bar{a}\,), \Lambda_1]$ and $S_T[y_*^{\perp}(\bar{a}\,), \Lambda_2]$, and $S_T[y_*^{\perp}(\bar{a}\,), \Psi_R]$ is independent of $S_T[y_*^{\perp}(\bar{a}\,), \Lambda_R]$.

Note we do not have a chi-square distributional result for $S_T[y_*^{\perp}(\bar{a}\,), \Lambda_3]$ which depends on the usual 2SLS residuals. On the other hand, $S_T[y_*^{\perp}(\bar{a}\,), \Lambda_4]$ follows a noncentral chi-square distribution, but it is not independent of $S_T[y_*^{\perp}(\bar{a}\,), \Psi_0]$.

The noncentrality parameters in Theorem 6.3 can be interpreted as *concentration parameters*. For example,

$$\begin{aligned}
\delta(\bar{a}, \Psi_0) &= T\,[\bar{\mu}_{y*}^{\perp}(\bar{a}\,)'\Psi_0 \bar{\mu}_{y*}^{\perp}(\bar{a}\,)] = T\,[\bar{\mu}_{y*}^{\perp}(\bar{a}\,)'C_1'\hat{\Delta}^{-1}C_1\bar{\mu}_{y*}^{\perp}(\bar{a}\,)] \\
&= \{M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}\}'C_1'(C_1C_1')^{-1}C_1\{M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}\} \\
&= \{M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}\}'\bar{\mathrm{P}}[C_1']\{M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}\}
\end{aligned} \tag{6.21}$$

and, under the simple complete linear model (2.7) - (2.8),

$$\delta(\bar{a}, \Psi_0) = (M_1 X_2 \Pi_2 \bar{a})'\bar{\mathrm{P}}[C_1'](M_1 X_2 \Pi_2 \bar{a}) = \bar{a}'\Pi_2' X_2' M_1 \bar{\mathrm{P}}[C_1'] M_1 X_2 \Pi_2 \bar{a}\,. \tag{6.22}$$

For $\delta(\bar{a}, \Psi_0)$ to be different from zero, we need $M_1 X_2 \Pi_2 \bar{a} \neq 0$. In particular, this requires that the instruments X_2 not be totally weak ($\Pi_2 \neq 0$) and linearly independent of X_1 ($M_1 X_2 \neq 0$). Similar interpretations can easily be formulated for the other centrality parameters. In particular, in the simple complete linear model, all noncentrality parameters are zero if $M_1 X_2 \Pi_2 \bar{a} = 0$. Note, however, this may not hold in the more general model described by (2.1) -(2.6), because of the nonlinear reduced form for Y and the presence of excluded instruments.

Theorem 6.3 allows us to conclude that $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{R}$ follow doubly noncentral F-distributions [*i.e.*, ratios of independent noncentral chi-square distributions] under the alternative hypothesis (conditional on $\bar{X}$ and V). This is spelled out in the following corollary.

Corollary 6.4 DOUBLY NONCENTRAL DISTRIBUTIONS FOR EXOGENEITY STATISTICS. *Under the model described by* (2.1) - (2.6)*, suppose Assumptions 2.1 and 2.2 hold. If* $\varepsilon \sim \mathrm{N}[0, I_T]$*, then conditional on* $\bar{X}$ *and* V*, we have :*

$$\mathscr{T}_1 \sim F[G, k_2 - G;\, \delta(\bar{a}, \Psi_0),\, \delta(\bar{a}, \Lambda_1)]\,, \tag{6.23}$$

$$\mathscr{T}_2 \sim F[G, T - k_1 - 2G;\, \delta(\bar{a}, \Psi_0),\, \delta(\bar{a}, \Lambda_2)]\,, \tag{6.24}$$

$$\mathscr{T}_4 = \frac{\kappa_4}{\kappa_2 \mathscr{T}_2^{-1} + 1} \leq \left(\frac{\kappa_4}{\kappa_2}\right) \mathscr{T}_2\,, \tag{6.25}$$

$$\mathscr{R} \sim F[k_2, T - k_1 - k_2 - G;\, \delta(\bar{a}, \Psi_R),\, \delta(\bar{a}, \Psi_R)]\,, \tag{6.26}$$

where the noncentrality parameters are defined in Theorem 6.3.

In the special case where (2.7) and (2.8) hold, we have $\Lambda_R M_1 g(X_1, X_2, X_3, \bar{\Pi}) = \Lambda_R g(X_1, X_2, X_3, \bar{\Pi}) = 0$ and $\delta(\bar{a}, \Psi_R) = 0$, so $\mathscr{R} \sim F[k_2, T - k_1 - k_2 - G;\, \delta(\bar{a}, \Psi_R)]$ the usual noncentral noncentral F-distribution. When $a = 0$, the distributions of $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{R}$ reduce to the central chi-square in (3.21) originally provided by Wu (1973) and Revankar and Hartley (1973). The setup under which these are obtained here is considerably more general than the usual linear reduced-form specification (2.7) considered by these authors.

Note $\mathscr{T}_4$ is proportional to a ratio of two noncentral chi-square distributions, but it is not doubly-noncentral chi-square due to the non-orthogonality of Ψ_0 and Λ_4 [$\Psi_0 \Lambda_4 = T^{-1}\Psi_0$, see (C.50)]. This observation carries to $\mathscr{H}_3$ through the identity $\mathscr{H}_3 = (T/\kappa_4)\mathscr{T}_4$. The same applies to $\mathscr{H}_1$ and $\mathscr{H}_2$, because of the presence of $S_T[y_*^{\perp}(\bar{a}\ ), \Lambda_3]$ in these statistics.

7. Simulation experiment

In this section, we study by simulation the finite-sample performance (size and power) of standard and exact MC DWH and RH tests. The DGP is described by equations (2.1) and (2.7) with $k_1 = 0$ (no X_1 matrix) and $Y = [Y_1, Y_2] \in \mathbb{R}^{T \times 2}$. The $T \times k_2$ instrument matrix X_2 was generated according to

$X_{2t} \overset{i.i.d.}{\sim} \mathbf{N}[0, I_{k_2}]$ for $t = 1, \ldots, T$, and kept fixed within the simulations. The value of $\beta = (\beta_1, \beta_2)'$ is set to $\beta_0 = (2, 5)'$; the results are qualitatively similar for alternative choices of β_0. The matrix Π_2 (which determines instrument strength) has the form $\Pi_2 = [\eta_1 \Pi_{01}, \eta_2 \Pi_{02}] \in \mathbb{R}^{k_2 \times 2}$ where $[\Pi_{01}, \Pi_{02}]$ is obtained by taking the first two columns of the identity matrix of order k_2.[6] η_1 and η_2 belong to the set $\{0, 0.01, 0.5\}$: $\eta_1 = \eta_2 = 0$ corresponds to complete identification failure, $\eta_1 = \eta_2 = 0.01$ is a design with weak identification, $\eta_1 \in \{0, 0.01\}$ and $\eta_2 = 0.5$ (or *vice versa*) can be interpreted as partial identification, and $\eta_1 = \eta_2 = 0.5$ corresponds to strong identification (strong instruments). These values entail the following theoretical R^2 coefficients for the (first-stage) regressions of Y_1 and Y_2 on X_2, respectively R_1^2 and R_2^2: (1) $R_1^2 = R_2^2 \in \{0, 0.0005, 0.56\}$ when $k_2 = 5$, and (2) $R_1^2 = R_2^2 \in \{0, 0.001, 0.71\}$ when $k_2 = 10$.[7] According to this measure, $R_1^2 = R_2^2 = 0$ represents complete identification failure, $R_1^2 = R_2^2 \in \{0, 0.0005, 0.001\}$ weak identification, $R_1^2 \in \{0, 0.0005, 0.001\}$ and $R_2^2 = 0.71$ [or $R_1^2 = 0.56$ and $R_2^2 \in \{0, 0.0005, 0.001\}$] partial identification, while $R_1^2 = 0.56$ and $R_2^2 = 0.71$ correspond to strong identification.

The errors u and V are generated so that

$$u = Va + e = V_1 a_1 + V_2 a_2 + e \tag{7.1}$$

where a_1 and a_2 are fixed scalar coefficients. In this experiment, we set $a = (a_1, a_2)' = \lambda\, a_0$, where $a_0 = (0.5, 0.2)'$ and $\lambda \in \{-5, -1, -0.1, 0, 0.1, 1, 5\}$; but the results do not change qualitatively with alternative values of a_0 and λ. In this setup, λ controls the endogeneity of $Y : \lambda = 0$ represents the null hypothesis of exogeneity (level), while values of λ different from zero represent the alternative hypothesis of endogeneity (power). We consider two specifications for the joint distribution of $[e, V]$. In the first one, $(e_t, V_t')' \overset{i.i.d.}{\sim} \mathbf{N}[0, I_3]$ for $t = 1, \ldots, T$ (Gaussian errors). In the second one, $e_t, V_{1t}, V_{2t}, t = 1, \ldots, T$ are i.i.d. according to a $t(3)$ distribution. MCT p-values are computed with $N = 199$ pseudo-samples. Except for the analysis of MC tests under misspecification of the error distribution, the sample size is $T = 50$. The simulations are based on 10000 replications, and the nominal level for the tests is 5%. Calculations were made using Matlab (2015b version).

7.1. Size and power of standard DWH tests

Table 1 presents the empirical rejection frequencies of standard DWH and RH tests with Gaussian errors (in which case exact critical values are available for $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{R}$). The first column of the table gives the different test statistics and the number of IVs (k_2), while the other columns report the rejection frequencies of the tests for each value of the endogeneity parameter (λ) and IV strength (η_1 and η_2).

The results confirm our theoretical analysis. The first half of Table 1 shows the results for $\eta_2 = 0$ (non-identification of β_2), the second part the results for $\eta_2 = 0.5$ (strong identification of β_2). In both cases, the identification of β_1 varies from completely weak to strong [$\eta_1 = 0, 0.01, 0.5$]. Our findings can be summarized as follows.

[6] We also ran the experiment where $[\Pi_{01}, \Pi_{02}]$ is the $k_2 \times 2$ matrix of ones, and found similar results to those presented.

[7] As $\mathbb{E}[X_{2t}] = 0$ and $\mathbb{E}[X_{2t} X_{2t}'] = I_{k_2}$ for all t, the R^2 of the first-stage regression of Y_j on X_2 is equal to $R_j^2 = k_2 \|\Pi_{2j}\|^2 (1 + k_2 \|\Pi_{2j}\|^2)^{-1} \equiv k_2 \eta_j^2 (1 + k_2 \eta_j^2)^{-1}$ for all $j \in \{1, 2\}$, where Π_{2j} is the jth column of Π_2.

Table 1. Size and power of standard exogeneity tests with Gaussian errors (nominal 5% level)

$\eta_2 = 0$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	4.9	5.1	63.6	4.7	5.0	19.5	4.6	5.1	5.1	5.0	4.8	5.5	5.2	5.1	5.0	5.1	4.8	5.5	5.4	4.8	5.6	5.2	4.9	19.5	4.9	4.9	63.3
	$\mathcal{T}_2$	4.7	5.1	100.0	5.1	4.7	51.4	4.9	5.1	5.4	5.4	5.1	5.0	4.5	5.0	5.0	4.9	5.0	5.2	5.3	5.1	5.3	5.0	4.8	51.4	4.6	5.3	100.0
	$\mathcal{T}_3$	0.0	0.0	97.1	0.0	0.0	27.6	0.0	0.0	0.9	0.0	0.0	0.6	0.0	0.0	0.8	0.0	0.0	0.7	0.0	0.0	0.8	0.0	0.0	27.3	0.0	0.0	97.1
$k_2 = 5$	$\mathcal{T}_4$	4.4	4.7	100.0	4.7	4.3	50.2	4.6	4.9	5.0	5.0	4.8	4.6	4.2	4.7	4.6	4.5	4.7	4.8	4.9	4.7	5.1	4.6	4.4	50.2	4.4	4.9	100.0
	$\mathcal{H}_1$	0.0	0.0	90.1	0.0	0.0	17.2	0.0	0.0	0.4	0.0	0.0	0.2	0.0	0.0	0.3	0.0	0.0	0.3	0.0	0.0	0.3	0.0	0.0	17.2	0.0	0.0	90.1
	$\mathcal{H}_2$	0.0	0.0	97.4	0.0	0.1	30.0	0.0	0.0	1.1	0.0	0.0	0.8	0.0	0.0	0.9	0.0	0.0	0.8	0.0	0.0	0.9	0.0	0.0	30.1	0.0	0.0	97.3
	$\mathcal{H}_3$	5.0	5.4	100.0	5.4	5.0	52.5	5.3	5.4	5.7	5.6	5.5	5.3	4.9	5.3	5.3	5.2	5.4	5.5	5.5	5.4	5.7	5.3	5.2	52.5	4.9	5.7	100.0
	$\mathcal{R}$	4.4	5.5	100.0	4.8	4.9	41.1	5.3	4.6	5.4	5.2	5.0	4.9	4.7	5.1	5.0	4.9	5.3	5.1	5.2	5.3	5.1	4.8	4.8	40.8	4.7	5.4	100.0
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	5.3	5.0	97.7	4.9	5.3	36.9	5.0	4.7	5.4	5.1	4.8	4.6	5.3	4.8	4.9	4.8	5.1	5.4	5.0	4.9	5.1	5.1	5.2	36.9	5.0	5.2	97.6
	$\mathcal{T}_2$	4.9	5.0	100.0	5.2	5.3	56.8	5.2	5.2	5.2	5.5	4.7	4.5	5.3	4.9	4.9	5.0	5.0	5.3	4.5	4.9	5.4	5.0	5.0	57.6	4.8	5.6	100.0
	$\mathcal{T}_3$	0.3	0.3	100.0	0.4	0.3	46.4	0.4	0.3	1.8	0.3	0.2	1.5	0.4	0.2	1.6	0.3	0.4	1.7	0.3	0.3	1.8	0.4	0.3	47.5	0.2	0.4	100.0
$k_2 = 10$	$\mathcal{T}_4$	4.7	4.6	100.0	4.8	5.0	55.6	4.8	4.9	5.1	5.1	4.4	4.3	4.9	4.5	4.6	4.8	4.7	5.0	4.2	4.6	5.0	4.7	4.7	56.4	4.5	5.1	100.0
	$\mathcal{H}_1$	0.1	0.1	98.5	0.2	0.2	30.8	0.1	0.1	0.6	0.2	0.1	0.5	0.2	0.1	0.6	0.1	0.2	0.7	0.2	0.2	0.7	0.1	0.2	31.2	0.1	0.3	98.6
	$\mathcal{H}_2$	0.4	0.4	100.0	0.5	0.4	48.9	0.5	0.4	2.2	0.4	0.3	1.8	0.4	0.3	1.9	0.4	0.5	2.1	0.4	0.4	2.2	0.5	0.4	50.1	0.3	0.5	100.0
	$\mathcal{H}_3$	5.2	5.2	100.0	5.6	5.5	57.9	5.3	5.5	5.6	5.7	5.1	4.8	5.6	5.1	5.3	5.2	5.3	5.6	4.7	5.3	5.8	5.3	5.4	58.7	5.1	5.9	100.0
	$\mathcal{R}$	4.8	5.5	100.0	4.7	5.3	37.0	5.0	4.8	5.0	5.0	4.9	4.9	4.8	4.9	4.7	4.7	5.7	5.6	4.8	5.1	5.0	5.0	4.8	37.1	4.8	5.7	100.0
$\eta_2 = 0.5$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	40.7	43.2	74.4	7.5	7.9	23.8	5.1	5.5	5.4	5.2	5.2	5.5	4.8	5.2	5.0	5.0	5.3	5.0	4.9	5.3	5.4	7.8	7.8	24.9	41.9	43.2	74.2
	$\mathcal{T}_2$	97.9	98.7	100.0	12.0	12.9	64.0	5.2	5.0	5.4	5.6	5.2	5.0	4.6	4.9	4.9	4.8	5.2	5.4	5.5	5.5	5.4	11.8	12.5	64.0	98.0	98.9	100.0
	$\mathcal{T}_3$	85.6	87.5	98.2	3.2	3.6	42.3	0.9	0.7	1.1	0.8	0.9	0.9	0.5	0.7	0.9	0.7	0.7	0.9	0.7	0.7	1.0	3.2	3.6	42.4	84.8	87.7	98.2
$k_2 = 5$	$\mathcal{T}_4$	97.8	98.6	100.0	11.3	12.2	62.8	4.6	4.6	5.1	5.3	4.8	4.7	4.5	4.6	4.6	4.5	4.8	5.0	5.2	5.1	5.1	11.3	11.9	62.6	97.9	98.7	100.0
	$\mathcal{H}_1$	70.7	72.8	89.4	1.6	1.8	25.3	0.5	0.3	0.3	0.3	0.5	0.4	0.3	0.2	0.4	0.3	0.3	0.3	0.3	0.3	0.3	1.5	1.7	24.6	68.7	73.1	89.9
	$\mathcal{H}_2$	87.0	88.6	98.3	3.8	4.3	45.0	1.0	0.7	1.4	1.0	1.1	1.1	0.7	0.9	1.1	0.8	0.8	1.1	0.9	0.9	1.1	3.7	4.3	44.8	86.3	88.9	98.4
	$\mathcal{H}_3$	98.0	98.8	100.0	12.6	13.5	65.0	5.4	5.2	5.7	5.9	5.6	5.3	4.9	5.1	5.1	5.1	5.5	5.7	5.8	5.9	5.8	12.3	13.0	65.1	98.1	98.9	100.0
	$\mathcal{R}$	96.2	97.8	100.0	9.6	9.8	51.3	5.2	4.8	5.5	5.1	5.0	5.0	4.5	5.1	4.8	4.8	5.2	5.1	5.3	5.2	5.4	8.8	9.8	51.2	96.2	97.7	100.0
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	78.3	80.7	99.0	10.2	10.9	42.2	4.9	4.9	5.3	4.7	4.7	4.7	5.2	4.8	4.9	4.8	5.1	5.2	5.0	5.3	5.3	10.4	11.5	43.4	77.3	81.5	99.1
	$\mathcal{T}_2$	98.5	98.9	100.0	13.2	13.5	64.7	5.0	4.9	5.2	4.9	4.8	4.6	5.3	4.8	5.0	5.1	5.0	5.2	4.9	5.2	5.3	13.0	14.2	65.6	98.5	99.2	100.0
	$\mathcal{T}_3$	97.1	97.9	100.0	7.0	7.5	55.6	1.8	1.7	2.0	1.4	1.4	1.6	1.7	1.6	1.6	1.8	1.6	1.7	1.5	1.8	1.8	6.9	8.0	56.7	97.1	98.2	100.0
$k_2 = 10$	$\mathcal{T}_4$	98.4	98.8	100.0	12.5	12.8	63.5	4.7	4.5	4.8	4.7	4.3	4.2	4.9	4.5	4.7	4.9	4.6	4.8	4.6	4.8	4.9	12.3	13.6	64.5	98.4	99.1	100.0
	$\mathcal{H}_1$	87.5	89.6	97.2	3.6	3.5	32.1	0.6	0.7	0.6	0.5	0.4	0.4	0.7	0.6	0.4	0.6	0.6	0.3	0.6	0.7	0.6	3.4	3.9	33.0	87.8	90.2	97.4
	$\mathcal{H}_2$	97.5	98.2	100.0	8.1	8.7	58.3	2.1	2.0	2.4	1.8	1.8	1.9	2.0	2.0	1.9	2.2	1.9	1.9	1.8	2.1	2.1	7.9	9.2	59.4	97.5	98.5	100.0
	$\mathcal{H}_3$	98.6	98.9	100.0	13.8	14.3	65.6	5.3	5.1	5.5	5.2	5.0	5.0	5.5	5.1	5.4	5.4	5.3	5.5	5.2	5.5	5.6	13.6	14.9	66.6	98.5	99.3	100.0
	$\mathcal{R}$	95.4	97.0	100.0	8.2	9.1	43.1	5.1	4.9	5.1	5.0	4.9	4.9	4.8	5.2	4.7	5.0	5.6	5.5	4.8	5.2	4.9	8.4	9.0	43.1	95.8	97.2	100.0

First, the rejection frequencies of all tests under the null hypothesis of exogeneity ($\lambda = 0$) are smaller or very close to the nominal 5% level, whether identification is weak ($\eta_2 = 0$ and $\eta_1 \in \{0, 0.01\}$), partial ($\eta_2 = 0.5$ and $\eta_1 \in \{0, 0.01\}$ or $\eta_1 = 0.5$ and $\eta_2 \in \{0, 0.01\}$), or strong ($\eta_1 = \eta_2 = 0.5$). Thus, all tests considered are valid in finite samples and robust to weak instruments (*i.e.*, the level is controlled). This confirms the analysis of Section 4. As expected, the tests $\mathscr{T}_2$, $\mathscr{T}_4$, $\mathscr{H}_3$, and $\mathscr{R}$ have rejections close to the 5% nominal level. Meanwhile, $\mathscr{T}_3$, $\mathscr{H}_1$ and $\mathscr{H}_2$ are highly conservative when identification is weak [column $\lambda = 0$ for $\eta_2 = 0$ and $\eta_1 \in \{0, 0.01\}$ of the table].

Second, when $\lambda \neq 0$, we see that all tests have power whether identification is strong [$\eta_1 = \eta_2 = 0.5$] or just partial [($\eta_2 = 0.5$ and $\eta_1 = 0, 0.01$) or ($\eta_2 = 0$ and $\eta_1 = 0.5$)]. Indeed, the rejection frequencies are equal or close to 100% when identification is strong ($\eta_1 = 0.5$ and $\eta_2 = 0.5$), despite the small sample size ($T = 50$). As expected, all tests have low power when all instruments are weak ($\lambda \neq 0$ and $\eta_2 = 0, \eta_1 \in \{0, 0.01\}$). In particular, the rejection frequencies are close to 5% when $\lambda \neq 0$ and $\eta_2 = 0,\ \eta_1 \in \{0, 0.01\}$. This confirms the results of Theorems 6.2 and 6.3. Moreover, the tests $\mathscr{T}_2$, $\mathscr{H}_3$, and $\mathscr{T}_4$ have the best powers, while $\mathscr{H}_1$ [the standard version of the Hausman (1978) test] exhibits the worst power performance. The latter observation is especially notable under partial identification; see the first part of Table 1 for $\eta_2 = 0$ and $\eta_1 = 0.5$.

7.2. Performance of the exact Monte Carlo tests

We will now examine the performance of the proposed exact MC exogeneity tests. To simplify the exposition, we focus on three setups : (i) the model is correctly specified; (ii) incomplete model (instrument exclusion); and (iii) the error distribution (used to implement the MC tests) is misspecified.

7.2.1. Size and power when the model is correctly specified

We first study the size and power of the exact MC tests when the model is correctly specified in the sense that (2.7) is well specified (no instrument exclusion and correct error distribution). To shorten the presentation, we give the results where the variables $e_t, V_{1t}, V_{2t}, t = 1, \ldots, T$, are i.i.d. according to a $t(3)$ distribution.[8] The results are presented in Table 2 and confirm the theoretical findings of Section 4.2.

First, the rejection frequencies under exogeneity ($\lambda = 0$) of all MC tests are close to 5%, whether identification is weak [$\eta_2 = 0$ and $\eta_1 \in \{0,\ 0.01\}$], partial [($\eta_2 = 0.5$ and $\eta_1 \in \{0,\ 0.01\}$) or ($\eta_2 = 0$ and $\eta_1 = 0.5$)], or strong [$\eta_1 = \eta_2 = 0.5$]. This represents a substantial improvement, especially for the standard $\mathscr{T}_3$, $\mathscr{H}_1$, $\mathscr{H}_2$ tests.

Second, with $\lambda \neq 0$ (endogeneity), all MC tests exhibit power as long as identification is not very weak. Interestingly, $\mathscr{T}_3$, $\mathscr{H}_2$, and even $\mathscr{H}_1$ become competitive in terms of power when the MCT approach is applied. $\mathscr{T}_3$ and $\mathscr{H}_2$ are especially attractive [see the results for $\lambda \in \{-5, -1, 1, 5\}$]. In many cases, the MCT based on the $\mathscr{R}$ statistic [Revankar and Hartley (1973)] is outperformed by the MC tests based on $\mathscr{T}_l$, $l = 2, 3, 4$ and $\mathscr{H}_i$, $i = 1, 2, 3$. It is worth noting that the weak performance of the MC test based on $\mathscr{R}$, is not due to the t-distributional form of the errors in Table 2.

[8]The results where e_t and V_{jt} ($j = 1, 2$) have $\chi^2(1)$, $\chi^2(2)$, or standard normal distributions are qualitatively the same as those presented here, thus are omitted.

Indeed, similar patterns are observed with normal errors, but these results are omitted to shorten the presentation. A similar behaviour is observed with the standard version of this statistic (see Table 1 with $k_2 = 10$, $\lambda \in \{-1, 1\}$ and $\eta_1 = 0.5$).

We will now analyze the size and power of the MC tests under two forms of model misspecification : reduced-form misspecification (instrument exclusion) and error-distribution misspecification.

7.2.2. Robustness to instrument exclusion

In this section, we investigate the impact of instrument exclusion on the size and power of the proposed MC exogeneity tests. To do this, we consider the case where the true DGP for (2.7) contains k_2 instruments but only $k_2 - 2$ instruments are used in the computation of the statistics (thus 2 IVs are missing).[9] To simplify, we conduct the simulation for the case in which the missing IVs are strong : $\eta_2 = 0.5$ and $\eta_1 \in \{0, 0.01, 0.5\}$. The results presented here cover both partial identification of $\beta = (\beta_1, \beta_2)'$ $[\eta_1 \in \{0, 0.01\}]$ and strong identification of β $[\eta_1 = 0.5]$; the results are similar in the (less relevant) case where the missing IVs are weak. The experiment is conducted with both normal errors (first part of Table 3) and $t(3)$ errors (second part of Table 3). The other parameters remain the same as in Tables 1 - 2. The results appear in Table 3 and the main findings are the following.

First, when $\lambda = 0$ (exogeneity), the rejection frequencies of all MC tests are close to 5%, whether identification is partial ($\eta_1 \in \{0, 0.01\}$) or strong ($\eta_1 = 0.5$), with or without Gaussian errors. This confirms the theoretical finding (in Sections 4.1 - 4.2) : MC exogeneity tests are robust to instrument exclusion (level is controlled).

Second, when $\lambda \neq 0$ (endogeneity) with $t(3)$ errors (second part of Table 2 versus second part of Table 3), we see that excluding informative instruments can lead to a power loss (as expected), especially under partial identification ($\eta_1 \in \{0, 0.01\}$); see the second parts of Tables 2 and 3. However, although instrument exclusion affects the power of the MC tests, it is clear from Table 3 that the tests still exhibit substantial power, especially when endogeneity is large ($\lambda \in \{-5, 5\}$) and identification is not too weak ($\eta_1 \in \{0.01, 0.5\}$). Overall, our results indicate that the MC tests perform quite well even when relevant instruments are omitted from the reduced-form specification (2.7).

7.2.3. Sensitivity to misspecification of the error distribution

We now examine the sensitivity of the proposed MC exogeneity tests to misspecification of the error distribution used to perform the MCT procedure. To shorten the presentation, we present the results where the true DGP in (2.1) and (2.7) is such that e_t, V_{1t}, V_{2t}, $t = 1, \ldots, T$, are i.i.d. according to a $t(3)$ distribution, but a normal distribution (which is incorrect) is used to perform the MC tests.[10] Again, we set $\eta_2 = 0.5$ and $\eta_1 = 0$, 0.01, 0.5. To see how the effect of a misspecification in the error distribution changes as the sample size increases, we consider three sample sizes ($T = 50$,

[9] These results are qualitatively the same when the number of missing IVs is increased. Consequently, we only present the main findings with 2 missing IVs.

[10] We also ran the experiment with $\chi^2(1)$ and $\chi^2(2)$ distributions. The results are similar with those reported here.

Table 2. Size and power of exact MC exogeneity tests with $t(3)$ errors (at nominal 5% level)

$\eta_2=0$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathscr{T}_1$	5.3	5.0	40.6	4.7	5.1	12.0	4.8	5.4	5.2	5.0	5.0	5.4	5.1	4.8	4.8	4.9	5.2	5.3	5.3	5.2	5.5	5.1	4.6	12.9	4.7	4.7	38.9
	$\mathscr{T}_2$	5.5	5.0	99.3	5.0	5.2	31.6	4.9	5.1	5.1	5.0	4.6	5.2	5.0	5.4	4.6	5.0	4.8	5.1	4.9	5.1	5.5	6.0	4.9	31.2	4.8	5.5	99.4
	$\mathscr{T}_3$	5.1	4.9	95.1	4.9	4.8	34	4.9	5.9	5.3	4.9	4.5	5	5.4	5.5	4.6	5.1	4.9	5.4	4.6	5.1	5.2	5.3	5.0	33.6	4.9	5.6	95.4
$k_2=5$	$\mathscr{T}_4$	5.5	5.0	99.3	5.0	5.2	31.6	4.9	5.1	5.1	5.0	4.6	5.2	5.0	5.4	4.6	5.0	4.8	5.1	4.9	5.1	5.5	6.0	4.9	31.2	4.8	5.5	99.4
	$\mathscr{H}_1$	5.2	4.9	93.5	5.0	4.8	33.5	5.0	5.8	5.4	4.9	4.6	4.9	5.3	5.4	4.7	5.1	4.9	5.1	4.6	5.2	5.2	5.3	5.1	33.0	5.0	5.6	94
	$\mathscr{H}_2$	5.1	4.9	95.1	4.9	4.8	34.0	4.9	5.9	5.3	4.9	4.5	5.0	5.4	5.5	4.6	5.1	4.9	5.4	4.6	5.1	5.2	5.3	5.0	33.6	4.9	5.6	95.4
	$\mathscr{H}_3$	5.5	5	99.3	5.0	5.2	31.6	4.9	5.1	5.1	5.0	4.6	5.2	5.0	5.4	4.6	5.0	4.8	5.1	4.9	5.1	5.5	6.0	4.9	31.2	4.8	5.5	99.4
	$\mathscr{R}$	5.4	5.4	99.5	5.0	5.1	25.8	5.3	5.1	5.0	4.8	4.8	4.3	5.4	4.6	5.4	5.3	4.6	4.8	4.5	5.0	5.9	5.1	4.7	25.7	5.2	5.4	99.5
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathscr{T}_1$	5.2	5.1	84.1	4.9	5.1	23.7	5	4.8	5.2	4.6	4.9	5.6	4.7	4.6	4.7	5.2	5.3	5.6	5.1	4.9	4.5	4.7	5.4	22.5	4.5	4.6	84.7
	$\mathscr{T}_2$	5.2	5.6	99.6	5.0	5.4	38.8	5.0	4.8	5.4	4.8	5.4	5.4	4.7	4.7	4.9	4.8	5.3	5.0	5.1	5.2	5.3	5.0	5.1	36.9	4.2	4.3	99.4
	$\mathscr{T}_3$	4.9	5.4	99.6	5.3	5.2	43.2	4.8	4.7	5.4	4.6	5.0	5.2	4.8	4.8	4.9	4.9	4.9	5.2	5.1	5.0	5.5	4.8	5.1	41.7	4.6	4.3	99.4
$k_2=10$	$\mathscr{T}_4$	5.2	5.6	99.6	5.0	5.4	38.8	5.0	4.8	5.4	4.8	5.4	5.4	4.7	4.7	4.9	4.8	5.3	5.0	5.1	5.2	5.3	5.0	5.1	36.9	4.2	4.3	99.4
	$\mathscr{H}_1$	4.9	5.4	99.4	5.3	5.2	42.5	4.8	4.7	5.7	4.7	5.0	5.2	4.8	4.8	4.9	4.9	4.9	5.2	5.1	5 .0	5.4	4.7	5.0	41	4.5	4.4	99.3
	$\mathscr{H}_2$	4.9	5.4	99.6	5.3	5.2	43.2	4.8	4.7	5.4	4.6	5.0	5.2	4.8	4.8	4.9	4.9	4.9	5.2	5.1	5.0	5.5	4.8	5.1	41.7	4.6	4.3	99.4
	$\mathscr{H}_3$	5.2	5.6	99.6	5.0	5.4	38.8	5.0	4.8	5.4	4.8	5.4	5.4	4.7	4.7	4.9	4.8	5.3	5.0	5.1	5.2	5.3	5.0	5.1	36.9	4.2	4.3	99.4
	$\mathscr{R}$	5	5.4	99.6	5.0	4.8	28.1	4.9	5.4	5.4	4.6	5.5	4.4	4.9	4.9	5.1	5.3	4.9	5.0	5.3	5.0	5.3	4.6	5.4	25.5	5.4	4.7	99.4
$\eta_2=0.5$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathscr{T}_1$	24.9	27.6	58.9	5.9	6.9	17.3	5.0	4.9	5.2	5.4	5.1	5.0	5.3	5.0	4.7	5.6	5.1	5.3	5.3	5.1	5.5	6.3	6.6	16.5	24.7	25.3	59.3
	$\mathscr{T}_2$	80.6	83.8	99.8	8.7	9.1	46.0	4.9	5.2	5.3	4.9	5.1	5.4	5.1	4.8	4.9	5.5	5.1	5.7	4.7	5.3	5.6	9.4	9.2	46.8	79.9	83.0	99.9
	$\mathscr{T}_3$	77.2	80.2	98.3	9.3	10.0	50.7	5.2	5.5	5.3	4.8	4.9	5.1	4.8	4.8	5.1	5.3	4.5	5.3	5.2	5.3	5.7	10.4	10.4	51.1	77.5	78.8	98.8
$k_2=5$	$\mathscr{T}_4$	80.6	83.8	99.8	8.7	9.1	46.0	4.9	5.2	5.3	4.9	5.1	5.4	5.1	4.8	4.9	5.5	5.1	5.7	4.7	5.3	5.6	9.4	9.2	46.8	79.9	83.0	99.9
	$\mathscr{H}_1$	74.9	77.8	96.3	9.1	10.1	48.0	5.2	5.4	5.3	4.8	4.8	5.3	4.8	4.8	5.1	5.3	4.8	5.2	5.1	5.1	5.6	10.1	10.3	48.1	74.9	76.8	97.0
	$\mathscr{H}_2$	77.2	80.2	98.3	9.3	10.0	50.7	5.2	5.5	5.3	4.8	4.9	5.1	4.8	4.8	5.1	5.3	4.5	5.3	5.2	5.3	5.7	10.4	10.4	51.1	77.5	78.8	98.8
	$\mathscr{H}_3$	80.6	83.8	99.8	8.7	9.1	46.0	4.9	5.2	5.3	4.9	5.1	5.4	5.1	4.8	4.9	5.5	5.1	5.7	4.7	5.3	5.6	9.4	9.2	46.8	79.9	83.0	99.9
	$\mathscr{R}$	77.8	81.0	99.7	7.6	7.7	37.7	5.3	5.0	5.0	4.6	4.9	4.4	5.6	4.6	5.3	5.5	4.9	4.5	4.6	5.1	6.1	7.5	7.7	37.5	77.1	81.6	99.8
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathscr{T}_1$	55.3	57.1	95.1	8.2	8.3	33.4	4.7	4.9	5.5	5.0	5.1	5.8	4.9	4.7	5.1	5.3	5.5	5.5	4.8	5.1	4.8	7.9	8.7	31.9	53.5	55.5	95.7
	$\mathscr{T}_2$	86.5	88.4	99.8	9.9	10.1	52.8	5.0	4.9	5.5	5.1	4.4	5.6	5.3	5.1	4.8	4.9	5.4	5.1	4.9	5.2	5.2	10.6	10.8	51.4	85.8	87.9	99.8
	$\mathscr{T}_3$	88.5	90.5	99.9	11.0	11.9	57.5	5.0	4.7	5.7	4.8	4.4	5.4	5.4	4.9	4.8	5.0	5.4	5.0	5.1	5.4	5.4	11.7	13.0	57.2	88.2	89.9	99.8
$k_2=10$	$\mathscr{T}_4$	86.5	88.4	99.8	9.9	10.1	52.8	5.0	4.9	5.5	5.1	4.4	5.6	5.3	5.1	4.8	4.9	5.4	5.1	4.9	5.2	5.2	10.6	10.8	51.4	85.8	87.9	99.8
	$\mathscr{H}_1$	87.4	89.6	99.7	11.1	12.0	54.9	5.1	4.8	5.7	4.8	4.4	5.3	5.1	4.8	5.1	5.1	5.4	5.1	5.2	5.2	5.5	11.7	12.7	55.4	86.8	88.9	99.6
	$\mathscr{H}_2$	88.5	90.5	99.9	11.0	11.9	57.5	5.0	4.7	5.7	4.8	4.4	5.4	5.4	4.9	4.8	5.0	5.4	5.0	5.1	5.4	5.4	11.7	13.0	57.2	88.2	89.9	99.8
	$\mathscr{H}_3$	86.5	88.4	99.8	9.9	10.1	52.8	5.0	4.9	5.5	5.1	4.4	5.6	5.3	5.1	4.8	4.9	5.4	5.1	4.9	5.2	5.2	10.6	10.8	51.4	85.8	87.9	99.8
	$\mathscr{R}$	79.4	82.8	99.7	7.5	7.9	35.0	4.7	5.3	5.5	4.6	5.3	4.6	5.0	4.9	5.1	5.3	5.2	4.8	5.5	5.3	5.3	7.4	8.3	34.3	80.8	81.9	99.6

Table 3. Exact MC exogeneity tests with excluded instruments (nominal 5% level)

Normal errors

		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	12.1	12.0	14.5	5.6	5.7	8.4	5.2	5.1	5.2	5.3	4.9	4.8	4.6	5.2	4.7	5.2	4.9	4.5	5.3	5.1	5.1	4.8	5.6	8.7	11.6	11.9	16.2
	$\mathcal{T}_2$	84.3	87.3	99.8	9.1	10.1	35.2	4.9	4.4	5.4	4.7	4.6	5.3	4.6	5.0	4.8	5.0	4.8	4.9	4.7	5.0	5.5	8.8	10.4	35.2	85.4	88.0	99.9
	$\mathcal{T}_3$	74.0	74.8	87.5	10.0	10.2	36.2	4.9	5.2	5.5	4.9	5.3	5.1	4.8	4.5	5.1	4.9	5.1	4.9	4.6	4.8	5.3	9.9	11.2	37.4	71.8	75.4	88.1
$k_2=5$	$\mathcal{T}_4$	84.3	87.3	99.8	9.1	10.1	35.2	4.9	4.4	5.4	4.7	4.6	5.3	4.6	5.0	4.8	5.0	4.8	4.9	4.7	5.0	5.5	8.8	10.4	35.2	85.4	88.0	99.9
	$\mathcal{H}_1$	69.2	69.8	80.7	9.9	9.8	33.4	4.9	5.1	5.2	5.1	5.5	5.0	4.6	4.6	4.8	4.9	5.2	5.1	4.6	4.9	5.2	9.7	11.0	34.7	67.0	70.3	81.6
	$\mathcal{H}_2$	74.0	74.8	87.5	10.0	10.2	36.2	4.9	5.2	5.5	4.9	5.3	5.1	4.8	4.5	5.1	4.9	5.1	4.9	4.6	4.8	5.3	9.9	11.2	37.4	71.8	75.4	88.1
	$\mathcal{H}_3$	84.3	87.3	99.8	9.1	10.1	35.2	4.9	4.4	5.4	4.7	4.6	5.3	4.6	5.0	4.8	5.0	4.8	4.9	4.7	5.0	5.5	8.8	10.4	35.2	85.4	88.0	99.9
	$\mathcal{R}$	81.8	84.7	99.7	8.4	9.0	31.0	5.2	4.7	5.4	4.8	5.0	5.3	4.7	5.3	5.0	5.1	5.2	4.5	4.9	5.0	5.7	8.5	9.2	30.9	82.2	84.8	99.7

		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	25.0	27.7	42.0	6.3	6.6	11.5	4.7	4.6	4.8	5.6	5.2	4.9	5.1	4.8	4.8	5.2	5.4	5.4	5.0	4.7	5.4	6.2	6.4	12.8	26.4	26.6	44.3
	$\mathcal{T}_2$	53.3	58.0	84.3	7.1	7.3	18.9	4.9	5.0	5.5	5.3	5.5	5.1	5.2	4.7	5.2	4.9	5.5	4.5	4.8	4.6	5.5	7.2	7.4	19.4	54.4	54.8	86.2
	$\mathcal{T}_3$	59.8	63.2	88.7	7.9	8.1	21.8	5.0	4.7	5.7	5.1	5.8	5.3	4.9	4.3	5.0	4.9	5.2	5.2	5.1	4.7	5.1	8.0	8.1	22.6	60.6	60.0	89.4
$k_2=10$	$\mathcal{T}_4$	53.3	58.0	84.3	7.1	7.3	18.9	4.9	5.0	5.5	5.3	5.5	5.1	5.2	4.7	5.2	4.9	5.5	4.5	4.8	4.6	5.5	7.2	7.4	19.4	54.4	54.8	86.2
	$\mathcal{H}_1$	58.7	61.8	86.5	7.9	8.3	21.7	4.9	4.7	5.7	5.3	5.6	5.3	4.9	4.2	5.1	4.9	5.4	5.1	5.3	4.4	4.8	7.8	8.2	22.0	58.9	59.4	87.2
	$\mathcal{H}_2$	59.8	63.2	88.7	7.9	8.1	21.8	5.0	4.7	5.7	5.1	5.8	5.3	4.9	4.3	5.0	4.9	5.2	5.2	5.1	4.7	5.1	8.0	8.1	22.6	60.6	60.0	89.4
	$\mathcal{H}_3$	53.3	58.0	84.3	7.1	7.3	18.9	4.9	5.0	5.5	5.3	5.5	5.1	5.2	4.7	5.2	4.9	5.5	4.5	4.8	4.6	5.5	7.2	7.4	19.4	54.4	54.8	86.2
	$\mathcal{R}$	41.0	46.2	74.1	6.1	6.2	12.6	4.6	4.5	5.1	5.3	4.9	4.6	5.1	5.1	4.9	4.3	5.2	4.3	4.5	4.6	4.8	5.9	6.6	13.0	43.0	43.5	74.7

$t(3)$ errors

		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	8.2	10.1	14.1	5.4	5.9	7.8	4.7	4.9	5.4	5.1	5.1	5.2	5.0	5.1	4.6	5.0	4.8	4.9	5.2	4.7	5.6	4.9	5.6	7.2	10.0	9.0	13.0
	$\mathcal{T}_2$	68.8	72.8	99.4	7.4	8.5	32.5	4.8	5.3	5.2	5.0	5.2	5.2	5.0	4.6	4.9	5.6	4.8	5.0	4.7	4.9	5.8	8.1	8.3	32.4	68.4	73.0	99.3
	$\mathcal{T}_3$	56.2	60.1	86.1	7.8	8.3	31.5	5.0	5.2	5.1	5.0	5.1	4.5	4.4	4.5	5.2	5.0	4.7	5.1	4.9	5.3	5.7	8.0	8.2	32.1	57.5	58.6	86.6
$k_2=5$	$\mathcal{T}_4$	68.8	72.8	99.4	7.4	8.5	32.5	4.8	5.3	5.2	5.0	5.2	5.2	5.0	4.6	4.9	5.6	4.8	5.0	4.7	4.9	5.8	8.1	8.3	32.4	68.4	73.0	99.3
	$\mathcal{H}_1$	53.7	57.5	81.9	7.8	8.3	30.2	5.1	5.1	5.2	5.0	5.0	4.5	4.4	4.6	5.2	4.9	4.8	5.1	4.8	5.3	5.7	8.0	8.2	30.8	54.7	56.8	82.2
	$\mathcal{H}_2$	56.2	60.1	86.1	7.8	8.3	31.5	5.0	5.2	5.1	5.0	5.1	4.5	4.4	4.5	5.2	5.0	4.7	5.1	4.9	5.3	5.7	8.0	8.2	32.1	57.5	58.6	86.6
	$\mathcal{H}_3$	68.8	72.8	99.4	7.4	8.5	32.5	4.8	5.3	5.2	5.0	5.2	5.2	5.0	4.6	4.9	5.6	4.8	5.0	4.7	4.9	5.8	8.1	8.3	32.4	68.4	73.0	99.3
	$\mathcal{R}$	67.3	70.7	99.3	7.4	7.8	29.3	5.1	5.3	5.6	4.7	4.9	4.6	5.0	4.5	5.7	5.5	5.0	5.5	4.6	5.1	5.8	7.4	7.8	29.4	67.4	71.9	99.3

		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	32.2	34.4	63.5	6.9	7.1	18.3	4.9	5.4	5.1	5.7	5.1	5.5	4.8	4.6	5.2	5.3	5.2	4.8	5.0	5.0	5.2	6.2	7.2	16.0	30.4	31.2	61.6
	$\mathcal{T}_2$	68.4	71.0	97.9	9.0	8.5	30.2	5.2	4.9	5.1	5.4	4.4	5.3	5.0	5.0	5.1	4.7	5.2	5.3	5.1	5.3	5.6	8.1	8.4	28.7	64.6	70.1	97.3
	$\mathcal{T}_3$	71.0	74.1	98.2	9.4	9.1	35.4	5.4	5.1	5.5	5.5	4.6	5.1	4.8	5.3	5.3	4.9	5.4	5.1	5.3	5.1	5.4	8.9	9.5	33.9	68.8	73.0	97.9
$k_2=10$	$\mathcal{T}_4$	68.4	71.0	97.9	9.0	8.5	30.2	5.2	4.9	5.1	5.4	4.4	5.3	5.0	5.0	5.1	4.7	5.2	5.3	5.1	5.3	5.6	8.1	8.4	28.7	64.6	70.1	97.3
	$\mathcal{H}_1$	69.2	72.6	96.8	9.2	9.2	33.9	5.2	5.1	5.5	5.4	4.5	4.9	4.8	5.3	5.3	4.7	5.4	5.0	5.3	5.2	5.3	8.9	9.8	32.7	67.5	71.5	96.2
	$\mathcal{H}_2$	71.0	74.1	98.2	9.4	9.1	35.4	5.4	5.1	5.5	5.5	4.6	5.1	4.8	5.3	5.3	4.9	5.4	5.1	5.3	5.1	5.4	8.9	9.5	33.9	68.8	73.0	97.9
	$\mathcal{H}_3$	68.4	71.0	97.9	9.0	8.5	30.2	5.2	4.9	5.1	5.4	4.4	5.3	5.0	5.0	5.1	4.7	5.2	5.3	5.1	5.3	5.6	8.1	8.4	28.7	64.6	70.1	97.3
	$\mathcal{R}$	58.9	63.3	95.8	7.1	6.8	20.7	4.8	5.2	5.9	4.2	4.9	4.8	4.9	4.6	5.0	5.0	5.6	5.3	5.6	5.3	5.2	7.4	7.7	19.8	60.1	61.5	95.9

100, 500). The other parameters remain the same as in Table 3. The results are presented in Table 4.

First, the rejection frequencies of all MC exogeneity tests are close to 5%, whether identification is partial ($\eta_1 = 0, 0.01$) or strong ($\eta_1 = 0.5$), for all sample sizes considered. The MCT method (approximately) achieves level control, even with a small sample size ($T = 50$).

Second, when $\lambda \neq 0$ (endogeneity), misspecification of the error distributions has little impact on the power of the MC tests (compare the first part of Table 4 with the second part of Table 2).

Third, as the sample T increases, the rejection frequencies of the MC exogeneity tests increase and almost reach 100% under strong endogeneity [see the columns for $\lambda = -5,\ -1,\ 1, 5$ and $T =$ 100, 500].

Overall, the MC exogeneity tests appear to be reasonably robust to a misspecification of the simulated error distribution, even with a small sample size.

8. Empirical illustrations

We illustrate our theoretical results on exogeneity tests through two empirical applications related to important issues in macroeconomics and labor economics literature : (1) the relation between trade and growth [Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw, Romer and Weil (1992)]; (2) the standard problem of measuring returns to education [Dufour and Taamouti (2007), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist, Imbens and Krueger (1999), Mankiw et al. (1992)].

8.1. Trade and growth

The trade and growth model studies the relationship between standards of living and openness. Frankel and Romer (1999) argued that trade share (ratio of imports or exports to GDP) which is the commonly used indicator of openness should be viewed as endogenous. So, instrumental variables method should be used to estimate the income-trade relationship.

The equation studied is

$$\ln(\text{Inc}_i) = \beta_0 + \beta_1 \text{Trade}_i + \gamma_1 \ln(\text{Pop}_i) + \gamma_2 \ln(\text{Area}_i) + u_i, \quad i = 1, \ldots, T \tag{8.1}$$

where Inc_i is the income per capita in country i, Trade_i is the trade share (measured as a ratio of imports and exports to GDP), Pop_i is the population of country i, and Area_i is country i area. The first stage model for the Trade variable is given by

$$\text{Trade}_i = a + bX_i + c_1 \ln(\text{Pop}_i) + c_2 \ln(\text{Area}_i) + V_i, \quad i = 1, \ldots, T \tag{8.2}$$

where X_i is an instrument constructed on the basis of geographic characteristics. In this paper, we use the sample of 150 countries and the data include for each country : the trade share in 1985, the area and population (1985), per capita income (1985), and the fitted trade share (instrument).

We wish to assess the exogeneity of the trade share variable in (8.1). The F-statistic in the first stage regression (8.2) is around 13 [see Frankel and Romer (1999, Table 2, p.385) and Dufour

Table 4. Size and power of MC tests under misspecification of the conditional distribution of the errors (level 5%) : $t(3)$ disturbances and Monte Carlo tests based on normal disturbances

$\eta_2=0.5$, $T=50$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	25.0	25.9	59.2	6.3	6.4	16.6	4.4	5.2	5.5	4.8	5.1	5.3	4.9	4.8	5.0	4.9	4.6	5.5	4.8	5.1	5.3	6.0	6.4	16.7	25.2	24.0	57.7
	$\mathcal{T}_2$	80.3	84.3	99.9	8.4	9.2	46.5	4.6	5.6	5.3	4.7	5.2	5.5	4.9	4.9	4.8	5.0	5.0	5.4	5.2	4.9	5.1	9.1	8.9	47.6	81.0	83.5	99.8
	$\mathcal{T}_3$	77.8	80.4	98.2	9.1	10.3	51.5	5.1	5.4	5.8	4.8	5.1	4.7	5.0	5.0	4.9	5.3	5.1	5.1	4.9	4.9	5.5	9.9	9.7	50.6	78.4	80.2	98.3
$k_2=5$	$\mathcal{T}_4$	80.3	84.3	99.9	8.4	9.2	46.5	4.6	5.6	5.3	4.7	5.2	5.5	4.9	4.9	4.8	5.0	5.0	5.4	5.2	4.9	5.1	9.1	8.9	47.6	81.0	83.5	99.8
	$\mathcal{H}_1$	75.3	78.1	96.3	8.9	10.2	48.3	5.0	5.6	5.9	4.7	5.0	4.8	4.9	4.9	4.8	5.3	5.1	5.2	4.9	5.0	5.6	9.6	9.6	48.3	75.7	78.0	96.4
	$\mathcal{H}_2$	77.8	80.4	98.2	9.1	10.3	51.5	5.1	5.4	5.8	4.8	5.1	4.7	5.0	5.0	4.9	5.3	5.1	5.1	4.9	4.9	5.5	9.9	9.7	50.6	78.4	80.2	98.3
	$\mathcal{H}_3$	80.3	84.3	99.9	8.4	9.2	46.5	4.6	5.6	5.3	4.7	5.2	5.5	4.9	4.9	4.8	5.0	5.0	5.4	5.2	4.9	5.1	9.1	8.9	47.6	81.0	83.5	99.8
	$\mathcal{R}$	76.6	81.3	99.8	7.6	8.6	37.2	5.2	4.9	5.4	5.3	4.9	5.0	5.0	5.2	4.5	4.8	5.6	5.0	5.1	4.8	5.6	7.8	8.2	37.9	77.8	81.2	99.8
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	54.1	56.7	95.1	8.3	8.9	34.8	4.7	5.3	5.3	5.0	5.0	5.1	5.3	5.0	4.4	5.7	4.5	4.6	5.5	4.9	4.9	7.5	9.0	32.9	56.9	58.0	95.4
	$\mathcal{T}_2$	86.8	89.2	99.8	10.0	11.9	53.9	4.8	5.3	5.5	5.2	5.3	4.6	5.5	4.7	4.5	5.2	4.6	4.7	5.8	4.9	5.2	9.4	11.1	52.7	87.4	89.5	99.8
	$\mathcal{T}_3$	89.0	91.0	99.8	11.6	12.6	59.6	5.1	5.1	5.8	4.8	5.3	4.5	5.1	5.1	4.6	5.4	4.8	4.9	5.4	4.6	5.7	10.9	11.5	58.9	89.4	91.6	99.8
$k_2=10$	$\mathcal{T}_4$	86.8	89.2	99.8	10.0	11.9	53.9	4.8	5.3	5.5	5.2	5.3	4.6	5.5	4.7	4.5	5.2	4.6	4.7	5.8	4.9	5.2	9.4	11.1	52.7	87.4	89.5	99.8
	$\mathcal{H}_1$	88.3	89.7	99.5	11.8	12.6	57.7	5.0	5.2	5.8	4.7	5.4	4.7	5.2	5.1	4.4	5.4	4.7	5.0	5.5	4.6	5.5	10.6	11.7	56.9	88.3	90.3	99.6
	$\mathcal{H}_2$	89.0	91.0	99.8	11.6	12.6	59.6	5.1	5.1	5.8	4.8	5.3	4.5	5.1	5.1	4.6	5.4	4.8	4.9	5.4	4.6	5.7	10.9	11.5	58.9	89.4	91.6	99.8
	$\mathcal{H}_3$	86.8	89.2	99.8	10.0	11.9	53.9	4.8	5.3	5.5	5.2	5.3	4.6	5.5	4.7	4.5	5.2	4.6	4.7	5.8	4.9	5.2	9.4	11.1	52.7	87.4	89.5	99.8
	$\mathcal{R}$	80.6	84.0	99.7	8.0	8.2	35.2	4.8	5.0	5.1	5.5	5.1	4.4	5.2	5.3	5.0	5.5	4.9	5.4	5.2	5.6	5.2	7.7	8.1	35.3	81.0	84.5	99.7
$\eta_2=0.5$, $T=100$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	50.1	51.4	88.3	8.5	8.6	34.8	4.5	4.7	5.5	5.1	5.0	5.3	4.8	5.1	5.5	4.9	5.0	4.9	5.2	5.0	5.1	8.7	8.4	30.6	51.4	49.9	87.3
	$\mathcal{T}_2$	98.1	98.4	100.0	14.4	16.7	79.7	5.2	4.5	5.7	4.9	4.5	5.4	5.4	5.3	5.0	5.0	5.0	5.2	4.8	5.0	5.8	14.7	15.6	80.8	98.1	98.4	100.0
	$\mathcal{T}_3$	96.0	96.4	99.6	16.8	18.5	81.7	4.7	5.1	6.1	4.9	4.7	4.9	5.4	5.3	4.7	5.4	5.3	4.9	5.2	5.1	6.5	17.0	18.3	82.7	96.0	96.4	99.6
$k_2=5$	$\mathcal{T}_4$	98.1	98.4	100.0	14.4	16.7	79.7	5.2	4.5	5.7	4.9	4.5	5.4	5.4	5.3	5.0	5.0	5.0	5.2	4.8	5.0	5.8	14.7	15.6	80.8	98.1	98.4	100.0
	$\mathcal{H}_1$	94.7	95.4	99.0	16.7	18.0	78.5	5.0	5.2	6.3	4.9	4.9	5.2	5.5	5.2	4.9	5.5	5.3	4.9	5.3	4.9	6.1	16.6	18.2	79.4	94.6	95.3	99.1
	$\mathcal{H}_2$	96.0	96.4	99.6	16.8	18.5	81.7	4.7	5.1	6.1	4.9	4.7	4.9	5.4	5.3	4.7	5.4	5.3	4.9	5.2	5.1	6.5	17.0	18.3	82.7	96.0	96.4	99.6
	$\mathcal{H}_3$	98.1	98.4	100.0	14.4	16.7	79.7	5.2	4.5	5.7	4.9	4.5	5.4	5.4	5.3	5.0	5.0	5.0	5.2	4.8	5.0	5.8	14.7	15.6	80.8	98.1	98.4	100.0
	$\mathcal{R}$	97.0	97.5	100.0	10.8	13.2	69.8	5.3	4.8	5.2	4.5	4.5	4.9	4.8	5.1	4.6	5.0	5.2	5.0	4.9	5.4	5.8	11.3	12.1	71.9	97.2	97.2	100.0
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	η_1	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	R_1^2	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	90.9	93.0	99.9	13.8	15.0	68.6	5.3	5.2	5.9	5.2	4.9	5.3	4.6	4.8	4.5	4.5	5.2	5.4	5.2	4.8	5.5	14.1	13.8	67.5	92.5	92.7	99.9
	$\mathcal{T}_2$	99.2	99.3	100.0	18.2	19.8	87.3	5.6	4.8	6.3	5.3	5.2	5.4	4.8	4.5	4.5	5.3	5.1	5.5	5.2	5.5	6.3	19.1	21.1	87.2	99.1	99.3	100.0
	$\mathcal{T}_3$	99.4	99.4	100.0	21.9	24.0	90.4	5.5	4.9	6.8	4.9	5.0	5.3	4.4	4.5	4.5	5.3	5.1	5.5	5.3	5.5	6.9	22.2	24.6	90.2	99.4	99.4	100.0
$k_2=10$	$\mathcal{T}_4$	99.2	99.3	100.0	18.2	19.8	87.3	5.6	4.8	6.3	5.3	5.2	5.4	4.8	4.5	4.5	5.3	5.1	5.5	5.2	5.5	6.3	19.1	21.1	87.2	99.1	99.3	100.0
	$\mathcal{H}_1$	99.3	99.2	100.0	21.8	23.6	89.3	5.5	4.8	6.7	4.7	5.1	5.3	4.7	4.5	4.5	5.3	5.1	5.3	5.3	5.6	6.8	21.9	24.2	89.4	99.2	99.3	100.0
	$\mathcal{H}_2$	99.4	99.4	100.0	21.9	24.0	90.4	5.5	4.9	6.8	4.9	5.0	5.3	4.4	4.5	4.5	5.3	5.1	5.5	5.3	5.5	6.9	22.2	24.6	90.2	99.4	99.4	100.0
	$\mathcal{H}_3$	99.2	99.3	100.0	18.2	19.8	87.3	5.6	4.8	6.3	5.3	5.2	5.4	4.8	4.5	4.5	5.3	5.1	5.5	5.2	5.5	6.3	19.1	21.1	87.2	99.1	99.3	100.0
	$\mathcal{R}$	98.0	98.4	100.0	11.9	11.8	71.3	5.0	5.0	5.5	4.8	5.2	5.3	4.9	4.8	4.8	5.3	4.4	5.5	4.8	4.7	5.3	11.4	13.0	70.9	97.7	98.1	100.0
η_2, $T=500$																												
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	$\eta_1\rightarrow$	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	$R_1^2\rightarrow$	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	0.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56	0	.0005	.56
	$\mathcal{T}_1$	98.4	98.5	100.0	23.6	26.2	89.1	4.9	5.2	6.8	4.5	5.4	4.5	5.0	5.8	5.2	5.4	5.0	4.6	5.0	5.5	6.3	24.4	25.8	90.6	97.9	98.3	100.0
	$\mathcal{T}_2$	100.0	100.0	100.0	56.8	59.5	99.9	5.6	5.9	9.9	5.3	5.0	5.1	4.6	4.9	5.3	5.4	4.6	4.9	5.3	5.2	9.1	57.8	61.7	99.8	100.0	100.0	100.0
	$\mathcal{T}_3$	99.9	99.9	100.0	62.2	66.6	99.5	5.2	6.1	11.2	5.3	5.4	5.2	4.8	5.1	5.6	5.0	4.9	4.8	5.5	5.7	10.9	62.2	66.6	99.6	99.9	99.9	100.0
$k_2=5$	$\mathcal{T}_4$	100.0	100.0	100.0	56.8	59.5	99.9	5.6	5.9	9.9	5.3	5.0	5.1	4.6	4.9	5.3	5.4	4.6	4.9	5.3	5.2	9.1	57.8	61.7	99.8	100.0	100.0	100.0
	$\mathcal{H}_1$	99.8	99.7	99.9	61.1	65.9	99.1	5.4	6.1	11.1	5.3	5.3	5.2	4.7	5.0	5.7	5.0	4.8	4.7	5.6	5.6	10.6	61.1	65.5	99.1	99.8	99.7	100.0
	$\mathcal{H}_2$	99.9	99.9	100.0	62.2	66.6	99.5	5.2	6.1	11.2	5.3	5.4	5.2	4.8	5.1	5.6	5.0	4.9	4.8	5.5	5.7	10.9	62.2	66.6	99.6	99.9	99.9	100.0
	$\mathcal{H}_3$	100.0	100.0	100.0	56.8	59.5	99.9	5.6	5.9	9.9	5.3	5.0	5.1	4.6	4.9	5.3	5.4	4.6	4.9	5.3	5.2	9.1	57.8	61.7	99.8	100.0	100.0	100.0
	$\mathcal{R}$	100.0	100.0	100.0	43.6	45.7	99.7	5.3	5.7	8.4	5.2	5.5	5.0	4.4	5.5	4.6	5.4	4.7	4.8	5.9	5.2	7.6	43.5	48.1	99.6	100.0	100.0	100.0
		$\lambda=-5$			$\lambda=-1$			$\lambda=-0.1$			$\lambda=-0.01$			$\lambda=0$			$\lambda=0.01$			$\lambda=0.1$			$\lambda=1$			$\lambda=5$		
	$\eta_1\rightarrow$	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5	0	0.01	0.5
	$R_1^2\rightarrow$	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71	0	.001	.71
	$\mathcal{T}_1$	100.0	100.0	100.0	61.0	65.6	99.8	5.4	5.4	9.7	5.1	5.5	5.8	5.6	5.3	5.3	4.7	5.0	5.0	5.7	5.7	9.2	62.5	65	99.8	100	100	100
	$\mathcal{T}_2$	100.0	100.0	100.0	78.3	82.6	100.0	5.2	5.7	12.2	4.9	5.1	5.7	5.4	5.3	5.3	4.5	4.8	5.2	6.3	6.1	11.8	81.1	82.7	99.9	100	100	100
	$\mathcal{T}_3$	100.0	100.0	100.0	83.7	86.7	100.0	5.8	6.1	14.0	5.2	5.2	5.5	5.0	5.3	5.1	4.4	4.7	5.2	6.3	6.1	13.9	85.6	86.9	99.9	100	100	100
$k_2=10$	$\mathcal{T}_4$	100.0	100.0	100.0	78.3	82.6	100.0	5.2	5.7	12.2	4.9	5.1	5.7	5.4	5.3	5.3	4.5	4.8	5.2	6.3	6.1	11.8	81.1	82.7	99.9	100	100	100
	$\mathcal{H}_1$	100.0	100.0	100.0	83.3	86.2	100.0	5.8	6.2	14.1	5.2	5.1	5.6	5.0	5.5	5.1	4.4	4.8	4.9	6.3	6.3	13.7	84.9	86.4	99.9	100	100	100
	$\mathcal{H}_2$	100.0	100.0	100.0	83.7	86.7	100.0	5.8	6.1	14.0	5.2	5.2	5.5	5.0	5.3	5.1	4.4	4.7	5.2	6.3	6.1	13.9	85.6	86.9	99.9	100	100	100
	$\mathcal{H}_3$	100.0	100.0	100.0	78.3	82.6	100.0	5.2	5.7	12.2	4.9	5.1	5.7	5.4	5.3	5.3	4.5	4.8	5.2	6.3	6.1	11.8	81.1	82.7	99.9	100	100	100
	$\mathcal{R}$	100.0	100.0	100.0	53.4	59.6	99.9	4.9	5.5	7.7	5.0	4.9	5.4	5.2	5.5	5.6	4.8	5.4	5.6	5.7	5.4	7.9	54.9	59.1	99.9	100	100	100

and Taamouti (2007)], so the fitted instrument X does not appear to be weak. Table 5 presents the p-values of the DWH and RH tests computed from the tabulated and exact MC p-values. The MC p-values are computed for Gaussian and $t(3)$ errors. Because the model contains one instrument and one (supposedly) endogenous variable, the statistic $\mathscr{T}_1$ is not well defined and is omitted.

Table 5. Exogeneity in trade and growth model

Statistics	Estimation	Standard p-value (%)	MC p-value (%) (Gaussian errors)	MC p-value (%) [$t(3)$-errors]
$\mathscr{R}$	3.9221	4.95	4.98	5.38
$\mathscr{H}_1$	2.3883	12.23	6.14	5.99
$\mathscr{H}_2$	2.4269	11.93	6.12	5.96
$\mathscr{H}_3$	3.9505	4.67	5.39	5.66
$\mathscr{T}_2$	3.9221	4.95	5.39	5.66
$\mathscr{T}_3$	2.3622	12.43	6.12	5.96
$\mathscr{T}_4$	3.8451	4.99	5.49	5.66

We note that the p-values based on the usual asymptotic distributions are close to the 5% nominal level for $\mathscr{H}_3$, $\mathscr{T}_2$, $\mathscr{T}_4$ and $\mathscr{R}$. So, there is evidence against the exogeneity of the trade share (at nominal level of 5%) when these statistics are applied. Meanwhile, the standard asymptotic p-values for $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{T}_3$ are relatively large (around 12%), but the corresponding MC p-values are similar to those of the other statistics (close to 5%). This discrepancy may be related to the fact that $\mathscr{H}_1$, $\mathscr{H}_2$, and $\mathscr{T}_3$ do not have standard asymptotic distributions [see Doko Tchatoka and Dufour (2016)].

In the end, the exact MC p-values for all test statistics are close to the 5% level, and indicate that there is evidence of trade share endogeneity in this model. This is supported by the relatively large discrepancy between the OLS estimate of β_1 (0.28) and the 2SLS estimate (2.03). Overall, our results underscore the importance of size correction through the exact MC procedures proposed.

8.2. Education and earnings

We now consider the well known example of estimating the return to education [see Angrist and Krueger (1991); Angrist and Krueger (1995); and Bound, Jaeger and Baker (1995)]. The equation studies is a relationship where the log-weekly earning (y) is explained by the number of years of education (E) and several other covariates (age, age squared, 10 dummies for birth of year) :

$$y = \beta_0 + \beta_1 E + \sum_{i=1}^{k_1} \gamma_i X_i + u. \tag{8.3}$$

In this model, β_1 measures the return to education. Because education can be viewed as endogenous, Angrist and Krueger (1991) used instrumental variables obtained by interacting quarter of birth with the year of birth (in this application, we use 40 dummies as instruments). The basic idea is that individuals born in the first quarter of the year start school at an older age, and can therefore drop

Table 6. Exogeneity in education and earning model

Statistics	Estimation	Standard p-value (%)	MC p-value (%) (Gaussian errors)	MC p-value (%) [$t(3)$-errors]
$\mathscr{R}$	0.68	93.99	49.91	49.93
$\mathscr{H}_1$	1.34	24.76	24.26	24.30
$\mathscr{H}_2$	1.34	24.76	24.26	24.30
$\mathscr{H}_3$	1.35	24.54	24.26	24.30
$\mathscr{T}_1$	2.04	16.11	22.49	22.99
$\mathscr{T}_2$	1.35	24.54	24.26	24.30
$\mathscr{T}_3$	1.35	22.48	24.26	24.30
$\mathscr{T}_4$	1.35	24.54	24.26	24.30

out after completing less schooling than individuals born near the end of the year. Consequently, individuals born at the beginning of the year are likely to earn less than those born during the rest of the year. The first stage model for E is then given by

$$E = \pi_0 + \sum_{i=1}^{k_2} \pi_i X_i + \sum_{i=1}^{k_1} \phi_i X_i + V \tag{8.4}$$

where X is the instrument matrix. It is well known that the instruments X constructed in this way are very weak and explains very little of the variation in education; see Bound et al. (1995). The data set consists of the 5% public-use sample of the 1980 US census for men born between 1930 and 1939. The sample contains 329 509 observations.

As in Section 8.2, we want to assess the exogeneity of education in (8.3) - (8.4). Table 6 shows the results of the tests with both the usual and exact MC p-values. As seen, the p-values of all tests are quite large, thus suggesting that there is little evidence against the exogeneity of the education variable, even at 15% nominal level. This means that either the education variable is effectively exogenous or the instruments used are very poor so that the power of the test is flat, as shown in Section 6. The latter scenario is highly plausible from the previous literature [for example, see Bound et al. (1995)]. This viewed is reinforced by the small discrepancy between the OLS estimate (0.07) and the 2SLS estimate (0.08) of β_1.

9. Conclusion

This paper develops a finite-sample theory of the distribution of 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.

Our study of the finite-sample distributions of the statistics under the null hypothesis shows that all tests are robust to weak instruments, missing instruments or misspecified reduced forms – in the sense that level is controlled. Indeed, we provided a general characterization of the structure of the test statistics which allows one to perform exact MC tests under general parametric distributional assumptions, which are in no way restricted to the Gaussian case, including heavy-tailed distributions without moments. The tests so obtained are exact even in cases where identification fails (or is weak) and conventional asymptotic theory breaks down.

After proving a general invariance property, we provided a characterization of the power of the tests that clearly exhibits the factors which determine power. We showed 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 typically have power as soon as we have one strong instrument. Consequently, exogeneity tests can detect an exogeneity problem even if not all model parameters are identified, provided at least some parameters are identifiable.

Though the exact distributional theory given in this paper requires relatively specific distributional assumptions, the "finite-sample" procedures provided remain asymptotically valid in the same way (in the sense that test level is controlled) under standard asymptotic assumptions. We study this problem extensively in a separate paper [Doko Tchatoka and Dufour (2016)]. Further, even if exogeneity hypotheses can have economic interest by themselves, we also show there how exogeneity tests can be fruitfully applied to build pretest estimators which generally dominate OLS and 2SLS estimators when the exogeneity of explanatory variables is in uncertain.

A. Wu and Hausman test statistics

We show here that Durbin-Wu statistics can be expressed in the same way as alternative Hausman statistics. The statistics $\mathscr{T}_l$, $l = 1, 2, 3, 4$ are defined in Wu (1973, eqs. (2.1), (2.18), (3.16), and (3.20)) as:

$$\begin{aligned}
\mathscr{T}_1 &= \kappa_1 \frac{Q^*}{Q_1}, \mathscr{T}_2 = \kappa_2 \frac{Q^*}{Q_2}, \mathscr{T}_3 = \kappa_3 \frac{Q^*}{Q_3}, \mathscr{T}_4 = \kappa_4 \frac{Q^*}{Q_4}, && \text{(A.1)}\\
Q^* &= (b_1 - b_2)' \left[(Y'A_2Y)^{-1} - (Y'A_1Y)^{-1}\right]^{-1} (b_1 - b_2), && \text{(A.2)}\\
Q_1 &= (y - Yb_2)'A_2(y - Yb_2), Q_2 = Q_4 - Q^*, && \text{(A.3)}\\
Q_4 &= (y - Yb_1)'A_1(y - Yb_1), Q_3 = (y - Yb_2)'A_1(y - Yb_2), && \text{(A.4)}\\
b_i &= (Y'A_iY)^{-1}Y'A_iy, i = 1, 2, A_1 = M_1, A_2 = M - M_1\,, && \text{(A.5)}
\end{aligned}$$

where b_1 is the ordinary least squares estimator of β, and b_2 is the instrumental variables method estimator of β. So, in our notations, $b_1 \equiv \hat{\beta}$ and $b_2 \equiv \tilde{\beta}$. From (3.8) - (3.13), we have:

$$\begin{aligned}
Q^* &= T\,(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\tilde{\sigma}^2(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_2^{-1}(\tilde{\beta} - \hat{\beta})\,, && \text{(A.6)}\\
Q_1 &= T\,\tilde{\sigma}_1^2, \quad Q_3 = T\,\tilde{\sigma}^2, \quad Q_4 = T\hat{\sigma}^2\,, && \text{(A.7)}\\
Q_2 &= Q_4 - Q^* = T\hat{\sigma}^2 - T\,(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\,\tilde{\sigma}_2^2. && \text{(A.8)}
\end{aligned}$$

Hence, we can write $\mathscr{T}_l$ as:

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

where κ_l, and $\tilde{\Sigma}_l$ are defined in (3.8) - (3.13).

To obtain (3.17), set $\mathscr{T}_0 = (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta})$. Then $\tilde{\sigma}_2^2 = \hat{\sigma}^2 - \mathscr{T}_0$, $\mathscr{T}_4 = \kappa_4\mathscr{T}_0/\hat{\sigma}^2$, and

$$\mathscr{T}_2 = \kappa_2 \frac{\mathscr{T}_0}{\tilde{\sigma}_2^2} = \kappa_2 \frac{\mathscr{T}_0}{\hat{\sigma}^2 - \mathscr{T}_0} = \kappa_2 \frac{(\mathscr{T}_0/\hat{\sigma}^2)}{1 - (\mathscr{T}_0/\hat{\sigma}^2)} = \kappa_2 \frac{(\mathscr{T}_4/\kappa_4)}{1 - (\mathscr{T}_4/\kappa_4)}\,, \tag{A.9}$$

hence

$$\frac{\mathscr{T}_4}{\kappa_4} = \frac{(\mathscr{T}_2/\kappa_2)}{(\mathscr{T}_2/\kappa_2) + 1} = \frac{\mathscr{T}_2}{\mathscr{T}_2 + \kappa_2} = \frac{1}{(\kappa_2/\mathscr{T}_2) + 1}\,. \tag{A.10}$$

In the sequel of this appendix, we shall use the following matrix formulas which are easily established by algebraic manipulations [on the invertibility of matrix differences, see Harville (1997, Theorem 18.2.4)].

Lemma A.1 DIFFERENCE OF MATRIX INVERSES. *Let A and B be two nonsingular $r \times r$ matrices.*

Then

$$
\begin{aligned}
A^{-1}-B^{-1} &= B^{-1}(B-A)A^{-1}=A^{-1}(B-A)B^{-1} \\
&= A^{-1}(A-AB^{-1}A)A^{-1}=B^{-1}(BA^{-1}B-B)B^{-1}. \qquad \text{(A.11)}
\end{aligned}
$$

Furthermore, $A^{-1}-B^{-1}$ is nonsingular if and only if $B-A$ is nonsingular. If $B-A$ is nonsingular, we have :

$$
\begin{aligned}
(A^{-1}-B^{-1})^{-1} &= A(B-A)^{-1}B=A-A(A-B)^{-1}A=A+A(B-A)^{-1}A=A[A^{-1}+(B-A)^{-1}]A \\
&= B(B-A)^{-1}A=B(B-A)^{-1}B-B=B[(B-A)^{-1}-B^{-1}]B \\
&= A(A-AB^{-1}A)^{-1}A=B(BA^{-1}B-B)^{-1}B. \qquad \text{(A.12)}
\end{aligned}
$$

It is easy to see from condition (2.6) that $\hat{\Omega}_{IV}$, $\hat{\Omega}_{LS}$ and $\hat{\Sigma}_V$ are nonsingular. On setting $A=\hat{\Omega}_{IV}$ and $B=\hat{\Omega}_{LS}$, we get :

$$
B-A=\hat{\Omega}_{LS}-\hat{\Omega}_{IV}=\frac{1}{T}Y'M_1Y-\frac{1}{T}Y'N_1Y=\frac{1}{T}Y'(M_1-N_1)Y=\frac{1}{T}Y'MY=\frac{1}{T}\hat{V}'\hat{V}=\hat{\Sigma}_V\,, \quad \text{(A.13)}
$$

so $\hat{\Omega}_{LS}-\hat{\Omega}_{IV}$ is nonsingular. By Lemma A.1, $\hat{\Delta}=\hat{\Omega}_{IV}^{-1}-\hat{\Omega}_{LS}^{-1}=A^{-1}-B^{-1}$ is also nonsingular, and

$$
\begin{aligned}
\hat{\Delta}^{-1} &= A+A(B-A)^{-1}A=\hat{\Omega}_{IV}+\hat{\Omega}_{IV}(\hat{\Omega}_{LS}-\hat{\Omega}_{IV})^{-1}\hat{\Omega}_{IV}=\hat{\Omega}_{IV}+\hat{\Omega}_{IV}\,\hat{\Sigma}_V^{-1}\,\hat{\Omega}_{IV} \\
&= \frac{1}{T}\left[Y'N_1Y+Y'N_1Y(Y'MY)^{-1}Y'N_1Y\right]=\frac{1}{T}Y'N_1\left[I_T+Y(Y'MY)^{-1}Y'\right]N_1Y\,. \qquad \text{(A.14)}
\end{aligned}
$$

From the above form, it is clear that $\hat{\Delta}^{-1}$ is positive definite. Note also that

$$
\begin{aligned}
\hat{\Delta}^{-1} &= B(B-A)^{-1}B-B=\hat{\Omega}_{LS}(\hat{\Omega}_{LS}-\hat{\Omega}_{IV})^{-1}\hat{\Omega}_{LS}-\hat{\Omega}_{LS}=\hat{\Omega}_{LS}\hat{\Sigma}_V^{-1}\,\hat{\Omega}_{LS}-\hat{\Omega}_{LS} \\
&= \frac{1}{T}[(Y'M_1Y)(Y'MY)^{-1}(Y'M_1Y)-(Y'M_1Y)]=\frac{1}{T}Y'M_1[Y(Y'MY)^{-1}Y'-I_T]M_1Y. \qquad \text{(A.15)}
\end{aligned}
$$

The latter shows that $\hat{\Delta}^{-1}$ only depends on the least-squares residuals M_1Y and MY.

B. Regression interpretation of DWH test statistics

Let us now consider the regressions (3.23) - (3.26). Using $Y=\hat{Y}+\hat{V}$, $\hat{Y}=X\hat{\Pi}$ and $\hat{\Pi}=(X'X)^{-1}X'Y$, we see that the 2SLS residual vector $\tilde{u}$ for model (2.1) based on the instrument matrix $X=[X_1,X_2]$ can be written as

$$
\begin{aligned}
\tilde{u} &= y-Y\tilde{\beta}-X_1\tilde{\gamma}=(y-\hat{Y}\tilde{\beta}-X_1\tilde{\gamma})-\hat{V}\tilde{\beta}=M_1(y-\hat{Y}\tilde{\beta})-\hat{V}\tilde{\beta} \\
&= M_1(y-\hat{Y}\tilde{\beta}-\hat{V}\tilde{\beta})=M_1(y-Y\tilde{\beta}) \qquad \text{(B.1)}
\end{aligned}
$$

where $\tilde{\beta}$ and $\tilde{\gamma}$ are the 2SLS estimators of β and γ, and the different sum-of-squares functions satisfy :

$$S(\hat{\theta}) = S_*(\hat{\theta}_*)\,, \quad \tilde{u}'\tilde{u} = S(\hat{\theta}^0) = S_*(\hat{\theta}^0_*) = \tilde{S}(\hat{\theta}^0_{**})\,, \quad \tilde{S}(\hat{\theta}_{**}) = (y - Y\tilde{\beta})'M(y - Y\tilde{\beta})\,, \tag{B.2}$$

$$S(\hat{\theta}^0) - S(\hat{\theta}) = S_*(\hat{\theta}^0_*) - S_*(\hat{\theta}_*)\,. \tag{B.3}$$

Let $R = \begin{bmatrix} 0 & 0 & I_G \end{bmatrix}$, and $R_* = \begin{bmatrix} I_G & 0 & -I_G \end{bmatrix}$, so that $Rb = a$ and $R_*\theta_* = \beta - a$. The null hypotheses $H_0 : a = 0$ and $H_0^* : \beta = b$ can thus be written as

$$H_0 : R\theta = 0\,, \quad H_0^* : R_*\theta_* = 0. \tag{B.4}$$

Further, $\hat{\theta}_* = [\tilde{\beta}', \tilde{\gamma}', \tilde{b}']'$ and $\hat{\theta}^0_* = [\hat{\beta}', \hat{\gamma}', \hat{\beta}']'$, where $\hat{\beta}$ and $\hat{\gamma}$ are the OLS estimators of β and γ based on the model (2.1), and

$$\begin{aligned}
R_*\hat{\theta} &= \begin{bmatrix} I_G & 0 & -I_G \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\gamma} \\ \tilde{b} \end{bmatrix} = \tilde{\beta} - \tilde{b}\,, && \text{(B.5)} \\
\hat{\theta}^0_* &= \hat{\theta}_* + (Z_*'Z_*)^{-1}R_*' \left[R_*(Z_*'Z_*)^{-1}R_*'\right]^{-1}(-R_*\hat{\theta}_*)\,, && \text{(B.6)} \\
S(\hat{\theta}^0_*) - S(\hat{\theta}_*) &= (\hat{\theta}^0_* - \hat{\theta}_*)'Z_*'Z_*(\hat{\theta}^0_* - \hat{\theta}_*) = (R_*\hat{\theta}_*)' \left[R_*(Z_*'Z_*)^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*)\,, && \text{(B.7)}
\end{aligned}$$

where $Z_* = [\hat{Y}, X_1, \hat{V}]$. On writing $Z_* = [\hat{X}_1, \hat{V}]$, where $\hat{X}_1 = [\hat{Y}, X_1]$, we get :

$$Z_*'Z_* = \begin{bmatrix} (\hat{X}_1'\hat{X}_1) & 0 \\ 0 & (\hat{V}'\hat{V}) \end{bmatrix}, \quad (Z_*'Z_*)^{-1} = \begin{bmatrix} (\hat{X}_1'\hat{X}_1)^{-1} & 0 \\ 0 & (\hat{V}'\hat{V})^{-1} \end{bmatrix}, \tag{B.8}$$

$$(\hat{X}_1'\hat{X}_1)^{-1} = \begin{bmatrix} \hat{Y}'\hat{Y} & \hat{Y}'X_1 \\ X_1'\hat{Y} & X_1'X_1 \end{bmatrix}^{-1} = \begin{bmatrix} W_{YY} & W_{Y1} \\ W_{1Y} & W_{11} \end{bmatrix}, \tag{B.9}$$

where $W_{YY} = \left[(\hat{Y}'\hat{Y}) - \hat{Y}'X_1(X_1'X_1)^{-1}X_1'\hat{Y}\right]^{-1} = \left[\hat{Y}'M_1\hat{Y}\right]^{-1} = [Y'(M_1 - M)Y]^{-1}$,

$$(Z_*'Z_*)^{-1}R_*' = \begin{bmatrix} W_{YY} & W_{Y1} & 0 \\ W_{1Y} & W_{11} & 0 \\ 0 & 0 & (\hat{V}'\hat{V})^{-1} \end{bmatrix} \begin{bmatrix} I_G \\ 0 \\ -I_G \end{bmatrix} = \begin{bmatrix} W_{YY} \\ W_{1Y} \\ -(\hat{V}'\hat{V})^{-1} \end{bmatrix}, \tag{B.10}$$

$$R_*(Z_*'Z_*)^{-1}R_*' = W_{YY} + (\hat{V}'\hat{V})^{-1}\,, \tag{B.11}$$

$$\hat{\theta}^0_* - \hat{\theta}_* = \begin{bmatrix} \hat{\beta} - \tilde{\beta} \\ \hat{\gamma} - \tilde{\gamma} \\ \hat{\beta} - \tilde{b} \end{bmatrix} = \begin{bmatrix} W_{YY} \\ W_{1Y} \\ -(\hat{V}'\hat{V})^{-1} \end{bmatrix} \left[W_{YY} + (\hat{V}'\hat{V})^{-1}\right]^{-1}(\tilde{b} - \tilde{\beta})\,. \tag{B.12}$$

>From the latter equation, we see that

$$\hat{\beta} - \tilde{\beta} = W_{YY}\left[W_{YY} + (\hat{V}'\hat{V})^{-1}\right]^{-1}(\tilde{b} - \tilde{\beta}) = W_{YY}\left[W_{YY} + (\hat{V}'\hat{V})^{-1}\right]^{-1}\tilde{a}\,, \tag{B.13}$$

where $\tilde{a} = \tilde{b} - \tilde{\beta}$ is the OLS estimate of a in (3.24). Hence, we have

$$\begin{aligned} \tilde{a} = \tilde{b} - \tilde{\beta} &= \left[W_{YY} + (\hat{V}'\hat{V})^{-1}\right] W_{YY}^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= \left\{[Y'(M_1 - M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\}[Y'(M_1 - M)Y](\hat{\beta} - \tilde{\beta}), \end{aligned} \tag{B.14}$$

which entails that

$$\begin{aligned} S(\hat{\theta}^0_*) - S(\hat{\theta}_*) &= (R_*\hat{\theta}_*)'\left[R_*(Z_*'Z_*)^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*) \\ &= (\tilde{b} - \tilde{\beta})'\left\{[Y'(M_1 - M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\}^{-1}(\tilde{b} - \tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})'[Y'(M_1 - M)Y]\left\{[Y'(M_1 - M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\}[Y'(M_1 - M)Y](\hat{\beta} - \tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})'W_{YY}^{-1}\left[W_{YY} + (Y'MY)^{-1}\right]W_{YY}^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})'W_{YY}^{-1}\left[W_{YY} + (Y'M_1Y - W_{YY}^{-1})^{-1}\right]W_{YY}^{-1}(\hat{\beta} - \tilde{\beta}). \end{aligned} \tag{B.15}$$

Using Lemma A.1 with $A = W_{YY}^{-1}$ and $B = Y'M_1Y$ in (B.15), we then get :

$$\begin{aligned} S(\hat{\theta}^0_*) - S(\hat{\theta}_*) &= (\hat{\beta} - \tilde{\beta})'W_{YY}^{-1}\left[W_{YY} + (Y'M_1Y - W_{YY}^{-1})^{-1}\right]W_{YY}^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})'A\left[A^{-1} + (B - A)^{-1}\right]A(\hat{\beta} - \tilde{\beta}) = (\hat{\beta} - \tilde{\beta})'(B^{-1} - A^{-1})^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})'\{[Y'(M_1 - M)Y]^{-1} - (Y'M_1Y)^{-1}\}^{-1}(\hat{\beta} - \tilde{\beta}) \\ &= T(\tilde{\beta} - \hat{\beta})'[\hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}]^{-1}(\tilde{\beta} - \hat{\beta}) = T(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) \end{aligned} \tag{B.16}$$

where $\hat{\Omega}_{IV} = \frac{1}{T}Y'(M_1 - M)Y$ and $\hat{\Omega}_{LS} = \frac{1}{T}Y'M_1Y$. Since we have $S_*(\hat{\theta}^0_*) - S_*(\hat{\theta}_*) = S(\hat{\theta}^0) - S(\hat{\theta})$, we get from (B.16), (3.13) and (3.31) :

$$S(\hat{\theta}) = S(\hat{\theta}^0) - [S_*(\hat{\theta}^0_*) - S_*(\hat{\theta}_*)] = S(\hat{\theta}^0) - T(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\,\tilde{\sigma}_2^2. \tag{B.17}$$

It is also clear from (3.13) and (3.31) that

$$S(\hat{\theta}^0) = T\hat{\sigma}^2,\; S_*(\hat{\theta}^0_*) = T\tilde{\sigma}^2. \tag{B.18}$$

Hence, except for $\mathscr{H}_1$, the other statistics can be expressed as :

$$\mathscr{H}_2 = T\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}^0_*)}\right), \mathscr{H}_3 = T\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta}^0)}\right), \tag{B.19}$$

$$\mathscr{T}_1 = \kappa_1\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}^0_*) - \tilde{S}(\hat{\theta}_{**})}\right) = \kappa_1\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{\tilde{S}(\hat{\theta}^0_{**}) - \tilde{S}(\hat{\theta}_{**})}\right), \tag{B.20}$$

$$\mathscr{T}_2 = \kappa_2\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta})}\right), \quad \mathscr{T}_3 = \kappa_3\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S_*(\hat{\theta}^0_*)}\right), \quad \mathscr{T}_4 = \kappa_4\left(\frac{S(\hat{\theta}^0) - S(\hat{\theta})}{S(\hat{\theta}^0)}\right), \tag{B.21}$$

$$\mathscr{R} = \kappa_R \left(\frac{\bar{S}(\check{\theta}^0) - \bar{S}(\check{\theta})}{\bar{S}(\check{\theta})} \right). \tag{B.22}$$

C. Proofs

To establish Proposition 4.1, it will be useful to state some basic identities for the different components of alternative exogeneity test statistics.

Lemma C.1 PROPERTIES OF EXOGENEITY STATISTICS COMPONENTS. *The random vectors and matrices in* (3.1) - (3.14) *satisfy the following identities : setting*

$$B_1 := (Y'M_1Y)^{-1}Y'M_1, \quad B_2 := (Y'N_1Y)^{-1}Y'N_1, \tag{C.1}$$

$$C_1 := B_2 - B_1, \quad \Psi_0 := C_1'\hat{\Delta}^{-1}C_1, \quad N_2 := I_T - M_1YB_2, \tag{C.2}$$

we have

$$B_1 M_1 = B_1, \quad B_2 M_1 = B_2 N_1 = B_2, \quad B_1 Y = B_2 Y = I_G, \tag{C.3}$$

$$C_1 Y = 0, \quad C_1 X_1 = 0, \quad C_1 \bar{\mathrm{P}}[M_1Y] = 0, \quad C_1 M_1 = C_1 \bar{\mathrm{M}}[M_1Y] = C_1, \tag{C.4}$$

$$M_1YA_1 = \bar{\mathrm{P}}[M_1Y], \quad M_1 \Psi_0 M_1 = M_1 \Psi_0 = \Psi_0 M_1 = \Psi_0, \tag{C.5}$$

$$M_1\Psi_R M_1 = \Psi_R, \quad M_1 \Lambda_R M_1 = M \Lambda_R M = \Lambda_R, \tag{C.6}$$

$$B_1 B_1' = B_1 B_2' = B_2 B_1' = \frac{1}{T}\hat{\Omega}_{LS}^{-1}, \quad B_2 B_2' = \frac{1}{T}\hat{\Omega}_{IV}^{-1}, \tag{C.7}$$

$$C_1C_1' = \frac{1}{T}\left(\hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}\right) = \frac{1}{T}\hat{\Delta}, \quad C_1\Psi_0 = \frac{1}{T}C_1, \quad \Psi_0\Psi_0 = \frac{1}{T}\Psi_0, \tag{C.8}$$

$$\tilde{\beta} - \hat{\beta} = (B_2 - B_1)y = C_1 y = C_1 (M_1 y), \tag{C.9}$$

$$(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = y' \Psi_0 y = (M_1 y)' \Psi_0 (M_1 y), \tag{C.10}$$

$$y - Y\hat{\beta} = [I_T - Y B_1]y, \quad y - Y\tilde{\beta} = [I_T - Y B_2]y, \tag{C.11}$$

$$\hat{u} = M_1(y - Y\hat{\beta}) = \bar{\mathrm{M}}[\bar{Y}]y = M_1\bar{\mathrm{M}}[M_1Y]y = \bar{\mathrm{M}}[M_1Y](M_1 y), \tag{C.12}$$

$$M(y - Y\hat{\beta}) = M\bar{\mathrm{M}}[M_1Y]y = M\bar{\mathrm{M}}[M_1Y](M_1 y), \tag{C.13}$$

$$\begin{aligned} N_1 (y - Y\tilde{\beta}) &= M_1 P(y - Y\tilde{\beta}) = M_1 \bar{\mathrm{M}}[M_1PY]Py = \bar{\mathrm{M}}[N_1Y]N_1 y \\ &= PM_1(y - Y\tilde{\beta}) = \bar{\mathrm{M}}[PM_1Y]P(M_1 y), \end{aligned} \tag{C.14}$$

$$\tilde{u} = M_1(y - Y\tilde{\beta}) = N_2(M_1 y), \quad M(y - Y\tilde{\beta}) = MN_2(M_1 y), \tag{C.15}$$

$$\tilde{\sigma}^2 = \frac{1}{T}(M_1 y)' N_2' N_2 (M_1 y), \tag{C.16}$$

$$\hat{\sigma}^2 = \frac{1}{T} y'\bar{\mathrm{M}}[\bar{Y}]y = \frac{1}{T} y'M_1\bar{\mathrm{M}}[M_1Y]y = \frac{1}{T}(M_1 y)'\bar{\mathrm{M}}[M_1Y](M_1 y), \tag{C.17}$$

$$\tilde{\sigma}_1^2 = \frac{1}{T} y' N_1 \bar{\mathrm{M}}[N_1 Y] N_1 y = \frac{1}{T} (M_1 y)' P \bar{\mathrm{M}}[P M_1 Y] P (M_1 y)\,, \tag{C.18}$$

$$\tilde{\sigma}_2^2 = (M_1 y)' \left\{ \frac{1}{T} \bar{\mathrm{M}}[M_1 Y] - \Psi_0 \right\} (M_1 y)\,, \tag{C.19}$$

$$y' \Psi_R y = \frac{1}{T} y' \bar{\mathrm{P}} \big[\bar{\mathrm{M}}[\bar{Y}] X_2 \big] \bar{\mathrm{M}}[\bar{Y}] y = \frac{1}{T} (M_1 y)' \bar{\mathrm{P}} \big[\bar{\mathrm{M}}[\bar{Y}] X_2 \big] (M_1 y)\,, \tag{C.20}$$

$$\hat{\sigma}_R^2 = \frac{1}{T} y' \bar{\mathrm{M}}[Z] y = \frac{1}{T} (M_1 y)' \bar{\mathrm{M}}[Z] (M_1 y)\,. \tag{C.21}$$

PROOF OF LEMMA C.1 Using the idempotence of M_1 and (3.15), we see that :

$$B_1 M_1 = (Y' M_1 Y)^{-1} Y' M_1 M_1 = (Y' M_1 Y)^{-1} Y' M_1 = B_1\,, \tag{C.22}$$

$$B_2 M_1 = [Y' N_1 Y]^{-1} Y' N_1 M_1 = [Y' N_1 Y]^{-1} Y' N_1 = B_2 = B_2 N_1 = B_2 (M_1 - M)\,, \tag{C.23}$$

$$M_1 Y A_1 = M_1 Y (Y' M_1 Y)^{-1} Y' M_1 = \bar{\mathrm{P}}(M_1 Y)\,, \tag{C.24}$$

$$C_1 M_1 = B_2 M_1 - B_1 M_1 = B_2 - B_1 = C_1\,, \quad C_1 X_1 = C_1 M_1 X_1 = 0\,, \tag{C.25}$$

$$B_1 Y = (Y' M_1 Y)^{-1} Y' M_1 Y = I_G = (Y' N_1 Y)^{-1} Y' N_1 Y = B_2 Y\,, \tag{C.26}$$

$$C_1 Y = B_2 Y - B_1 Y = 0\,, \tag{C.27}$$

$$\begin{aligned} C_1 \bar{\mathrm{P}}[M_1 Y] &= [(Y' N_1 Y)^{-1} Y' N_1 - (Y' M_1 Y)^{-1} Y' M_1] M_1 Y (Y' M_1 Y)^{-1} Y' M_1 \\ &= [(Y' N_1 Y)^{-1} Y' N_1 Y - (Y' M_1 Y)^{-1} Y' M_1 Y] (Y' M_1 Y)^{-1} Y' M_1 \\ &= (I_G - I_G)(Y' M_1 Y)^{-1} Y' M_1 = 0\,, \end{aligned} \tag{C.28}$$

$$C_1 \bar{\mathrm{M}}[M_1 Y] = C_1 [I_T - \bar{\mathrm{P}}[M_1 Y] = C_1\,, \tag{C.29}$$

$$M_1 \bar{\mathrm{M}}[\bar{Y}] M_1 = \bar{\mathrm{M}}[\bar{Y}]\,, \quad M_1 \bar{\mathrm{M}}[Z] M_1 = \bar{\mathrm{M}}[Z]\,, \tag{C.30}$$

$$M_1 \Psi_R M_1 = \frac{1}{T} \{ M_1 \bar{\mathrm{M}}[\bar{Y}] M_1 - M_1 \bar{\mathrm{M}}[Z] M_1 \} = \Psi_R\,, \quad M_1 \Lambda_R M_1 = \frac{1}{T} M_1 \bar{\mathrm{M}}[Z] M_1 = \Lambda_R\,, \tag{C.31}$$

so (C.3) - (C.6) are established. (C.7) and (C.8) follow directly from (3.15) and the definitions of B_1, B_2, C_1 and Ψ_0. We get (C.9) and (C.10) by using the definitions of $\hat{\beta}$ and $\tilde{\beta}$ in (3.4) - (3.5). (C.11) follows on using (3.4) and (3.5). (C.12) comes from the fact that the residuals $M_1(y - Y\hat{\beta})$ are obtained by minimizing $\|y - Y\hat{\beta} - X_1\gamma\|^2$ with respect to γ, or equivalently $\|y - Y\beta - X_1\gamma\|^2$ with respect to β and γ. (C.13) follows from (C.12) and noting that $M = M M_1$. Similarly, the first identity in (C.14) comes from the fact that the residuals $M_1 P(y - Y\tilde{\beta}) = M_1(y - PY\tilde{\beta})$ are obtained by minimizing $\|y - PY\tilde{\beta} - X_1\gamma\|^2$ with respect to γ, or equivalently by minimizing $\|y - PY\beta - X_1\gamma\|^2$ with respect to β and γ. The others follow on noting that $N_1 = M_1 P = P M_1$ and

$$M_1 \bar{\mathrm{M}}[M_1 P Y] P = \bar{\mathrm{M}}[P M_1 Y] M_1 P = \bar{\mathrm{M}}[P M_1 Y] P M_1\,. \tag{C.32}$$

To get (C.15) and (C.16), we note that

$$\tilde{u} = y - Y\tilde{\beta} - X_1\tilde{\gamma} = M_1(y - Y\tilde{\beta}) = M_1\,[I_T - YA_2]y = [I_T - M_1YA_2](M_1\,y) = N_2(M_1\,y) \tag{C.33}$$

hence

$$\tilde{\sigma}^2 = \frac{1}{T}\tilde{u}'\tilde{u} = \frac{1}{T}(y - Y\tilde{\beta})'M_1M_1(y - Y\tilde{\beta}) = \frac{1}{T}(M_1\,y)'N_2'N_2(M_1\,y)\,. \tag{C.34}$$

Further, using (3.11) - (3.3), (C.12) and (C.14), we see that :

$$\hat{\sigma}^2 = \frac{1}{T}(y - Y\hat{\beta})'M_1(y - Y\hat{\beta}) = \frac{1}{T}y'\bar{\mathrm{M}}[\bar{Y}]y = \frac{1}{T}y'M_1\,\bar{\mathrm{M}}[M_1Y]\,y = \frac{1}{T}(M_1\,y)'\bar{\mathrm{M}}[M_1Y](M_1\,y)\,, \tag{C.35}$$

$$\begin{aligned}\tilde{\sigma}_1^2 &= \frac{1}{T}(y - Y\tilde{\beta})'N_1(y - Y\tilde{\beta}) = \frac{1}{T}(y - Y\tilde{\beta})'PM_1\,P(y - Y\tilde{\beta}) \\ &= \frac{1}{T}y'N_1'\bar{\mathrm{M}}[N_1Y]N_1\,y = \frac{1}{T}(M_1\,y)'P\bar{\mathrm{M}}[P\,M_1Y]P\,(M_1\,y)\,,\end{aligned} \tag{C.36}$$

$$\begin{aligned}\tilde{\sigma}_2^2 &= \hat{\sigma}^2 - (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = \frac{1}{T}\{y'M_1\,\bar{\mathrm{M}}[M_1Y]\,y\} - y'\,\Psi_0\,y \\ &= (M_1\,y)'\left\{\frac{1}{T}\bar{\mathrm{M}}[M_1Y] - \Psi_0\right\}(M_1\,y)\,,\end{aligned} \tag{C.37}$$

so (3.11) - (3.13) are established. Finally, (C.20) and (C.21) follow by observing that $M_1\bar{M}[\bar{Y}] = \bar{M}[\bar{Y}]M_1 = \bar{M}[\bar{Y}]M_1$ and $M_1\bar{M}[Z] = M_1\bar{M}[Z] = \bar{M}[Z]$, so that $M_1\bar{P}\big[\bar{M}[\bar{Y}]X_2\big]M_1 = \bar{P}\big[\bar{M}[\bar{Y}]X_2\big]$ and $M_1\bar{M}[Z]M_1 = \bar{M}[Z]$. □

Using Lemma C.1, we can now prove Proposition 4.1.

PROOF OF PROPOSITION 4.1 We first note that

$$\tilde{\beta} - \hat{\beta} = (B_2 - B_1)\,y = C_1\,y\,, \tag{C.38}$$

$$(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = y'\,C_1'\,\hat{\Delta}^{-1}C_1\,y = y'\,\Psi_0\,y\,, \tag{C.39}$$

so that, by the definitions (3.1) - (3.3),

$$\mathscr{T}_l = \kappa_l(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta} - \hat{\beta}) = \kappa_l\,\frac{(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta})}{\tilde{\sigma}_l^2} = \frac{y'\,\Psi_0\,y}{\tilde{\sigma}_l^2}\,, \quad l = 1, 2, 3, 4, \tag{C.40}$$

$$\mathscr{H}_i = T\,(\tilde{\beta} - \hat{\beta})'\hat{\Sigma}_i^{-1}(\tilde{\beta} - \hat{\beta}) = T\,\frac{(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta})}{\hat{\sigma}_i^2} = \frac{y'\,\Psi_0\,y}{\hat{\sigma}_i^2}\,, \quad i = 2, 3, \tag{C.41}$$

where, using Lemma C.1,

$$\tilde{\sigma}_1^2 = \frac{1}{T}(y - Y\tilde{\beta})' N_1 (y - Y\tilde{\beta}) = \frac{1}{T} y' N_1 \bar{\mathrm{M}}[N_1 Y] N_1 y = y' \Lambda_1 y\,, \tag{C.42}$$

$$\tilde{\sigma}_2^2 = y' M_1 \left\{ \frac{1}{T} \bar{\mathrm{M}}[M_1 Y] - \Psi_0 \right\} (M_1 y) = y' \Lambda_2 y\,, \tag{C.43}$$

$$\tilde{\sigma}_3^2 = \tilde{\sigma}^2 = \frac{1}{T} y' M_1 N_2' N_2 M_1 y = y' \Lambda_3 y\,, \tag{C.44}$$

$$\tilde{\sigma}_4^2 = \hat{\sigma}^2 = \frac{1}{T} y' \bar{\mathrm{M}}[\bar{Y}] y = \frac{1}{T} y' M_1 \bar{\mathrm{M}}[M_1 Y] M_1 y = y' \Lambda_4 y\,, \tag{C.45}$$

$$\hat{\sigma}_2^2 = \tilde{\sigma}^2 = y' \Lambda_3 y\,, \quad \hat{\sigma}_3^2 = \hat{\sigma}^2 = y' \Lambda_4 y\,. \tag{C.46}$$

For $\mathscr{H}_1$, we have

$$\mathscr{H}_1 = T(\tilde{\beta} - \hat{\beta})' \hat{\Sigma}_1^{-1} (\tilde{\beta} - \hat{\beta}) = T\, y' C_1' \hat{\Sigma}_1^{-1} C_1 y = T\,(y' \Psi_1[y]\, y) \tag{C.47}$$

where

$$\hat{\Sigma}_1 = \tilde{\sigma}^2 \hat{\Omega}_{IV}^{-1} - \hat{\sigma}^2 \hat{\Omega}_{LS}^{-1} = (y' \Lambda_3 y)\, \hat{\Omega}_{IV}^{-1} - (y' \Lambda_4 y)\, \hat{\Omega}_{LS}^{-1}\,. \tag{C.48}$$

The result for R follows directly by using (3.3). □

In order to characterize the null distributions of the test statistics (Theorem 4.2), it will be useful to first spell out some algebraic properties of the weighting matrices in Proposition 4.1. This is done by the following lemma.

Lemma C.2 PROPERTIES OF WEIGHTING MATRICES IN EXOGENEITY STATISTICS. *The matrices Ψ_0, Λ_1, Λ_2, Λ_4, Ψ_R and Λ_R in* (4.1) - (4.6) *satisfy the following identities :*

$$\Lambda_2 = \Lambda_4 - \Psi_0\,,\ C_1 \Lambda_1 = C_1 \Lambda_2 = \Psi_0 \Lambda_1 = \Psi_0 \Lambda_2 = \Psi_R \Lambda_R = 0\,, \tag{C.49}$$

$$C_1 \Lambda_4 = \frac{1}{T} C_1\,,\ \Psi_0 \Lambda_4 = \frac{1}{T} \Psi_0\,, \tag{C.50}$$

$$M_1 \Lambda_l M_1 = \Lambda_l\,,\ l = 1, \ldots, 4\,. \tag{C.51}$$

Further, the matrices $T\Psi_0$, $T\Lambda_1$, $T\Lambda_2$, $T\Lambda_4$, $T\Psi_R$ and $T\Lambda_R$ are symmetric idempotent.

PROOF OF LEMMA C.2 To get (C.49) - (C.50), we observe that :

$$\Lambda_2 = M_1 \left(\frac{1}{T} \bar{\mathrm{M}}[M_1 Y] - \Psi_0 \right) M_1 = \Lambda_4 - M_1 \Psi_0 M_1 = \Lambda_4 - \Psi_0\,, \tag{C.52}$$

$$C_1 N_1 \bar{\mathrm{P}}[N_1 Y] \quad = \quad \frac{1}{T}[B_2 - B_1] N_1 N_1 Y\, \hat{\Omega}_{IV}^{-1} Y' N_1 = \frac{1}{T}[\hat{\Omega}_{IV}^{-1} Y' N_1 - \hat{\Omega}_{LS}^{-1} Y' M_1] N_1 Y\, \hat{\Omega}_{IV}^{-1} Y' N_1$$

$$
\begin{aligned}
&= \frac{1}{T}[\hat{\Omega}_{IV}^{-1}Y'N_1Y\,\hat{\Omega}_{IV}^{-1}Y' - \hat{\Omega}_{LS}^{-1}Y'N_1Y\,\hat{\Omega}_{IV}^{-1}Y']N_1 = \frac{1}{T}[\hat{\Omega}_{IV}^{-1}Y' - \hat{\Omega}_{LS}^{-1}Y']N_1 \\
&= \frac{1}{T}[\hat{\Omega}_{IV}^{-1}Y'N_1 - \hat{\Omega}_{LS}^{-1}Y'M_1]N_1 = [B_2 - B_1]N_1 = C_1N_1\,, \qquad \text{(C.53)}
\end{aligned}
$$

$$
C_1M_1\,\bar{\mathrm{P}}[M_1Y] = C_1M_1Y(Y'M_1Y)^{-1}Y'M_1 = 0\,, \qquad \text{(C.54)}
$$

$$
\bar{\mathrm{M}}[\bar{Y}]\bar{\mathrm{M}}[Z] = \bar{\mathrm{M}}[Z]\,, \qquad \text{(C.55)}
$$

hence

$$
C_1\Lambda_1 = C_1\left(\frac{1}{T}N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1\right) = \frac{1}{T}C_1N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1 = \frac{1}{T}C_1N_1\left(I_T - \bar{\mathrm{P}}[N_1Y]\right)N_1 = 0\,, \qquad \text{(C.56)}
$$

$$
\begin{aligned}
C_1\Lambda_2 &= C_1M_1\left(\frac{1}{T}\bar{\mathrm{M}}[M_1Y] - \Psi_0\right)M_1 = \frac{1}{T}C_1M_1\bar{\mathrm{M}}[M_1Y]M_1 - C_1M_1\Psi_0M_1 \\
&= \frac{1}{T}C_1M_1\big(I_T - \bar{\mathrm{P}}[M_1Y]\big)M_1 - C_1\Psi_0 = \frac{1}{T}C_1 - \frac{1}{T}C_1 = 0\,, \qquad \text{(C.57)}
\end{aligned}
$$

$$
C_1\Lambda_4 = \frac{1}{T}C_1M_1\bar{\mathrm{M}}[M_1Y]M_1 = \frac{1}{T}C_1M_1\bar{\mathrm{M}}[M_1Y] = \frac{1}{T}C_1\,, \qquad \text{(C.58)}
$$

$$
\Psi_0\Lambda_4 = \frac{1}{T}C_1'\hat{\Delta}^{-1}C_1\,M_1\bar{\mathrm{M}}[M_1Y]M_1 = \frac{1}{T}C_1'\hat{\Delta}^{-1}C_1\,M_1\bar{\mathrm{M}}[M_1Y] = \frac{1}{T}C_1'\hat{\Delta}^{-1}C_1 = \frac{1}{T}\Psi_0\,, \qquad \text{(C.59)}
$$

$$
\Psi_0\Lambda_2 = \Psi_0M_1\left(\frac{1}{T}\bar{\mathrm{M}}[M_1Y] - \Psi_0\right)M_1 = \Psi_0\,(\Lambda_4 - \Psi_0) = \frac{1}{T}\Psi_0 - \frac{1}{T}\Psi_0 = 0\,, \qquad \text{(C.60)}
$$

$$
\Psi_R\Lambda_R = \frac{1}{T^2}\{\bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z]\}\bar{\mathrm{M}}[Z] = 0\,. \qquad \text{(C.61)}
$$

(C.51) follow directly from the idempotence of M_1 and the definitions of Λ_l, $l = 1, \ldots, 4$. Finally, the idempotence and symmetry of the weight matrices can be checked as follows:

$$
\begin{aligned}
(T\Psi_0)(T\Psi_0) &= T\,C_1'\hat{\Delta}^{-1}C_1C_1'\hat{\Delta}^{-1}C_1 = T^2\,C_1'\hat{\Delta}^{-1}\left(\frac{1}{T}\hat{\Delta}\right)\hat{\Delta}^{-1}C_1 = T\,C_1'\hat{\Delta}^{-1}C_1 \\
&= T\Psi_0 = T\Psi_0'\,, \qquad \text{(C.62)}
\end{aligned}
$$

$$
(T\Lambda_1)(T\Lambda_1) = \big(N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1\big)\big(N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1\big) = N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1 = T\Lambda_1 = T\Lambda_1'\,, \qquad \text{(C.63)}
$$

$$
(T\Lambda_4)(T\Lambda_4) = M_1\bar{\mathrm{M}}[M_1Y]M_1\,M_1\,\bar{\mathrm{M}}[M_1Y]M_1 = M_1\bar{\mathrm{M}}[M_1Y]M_1 = T\Lambda_4 = T\Lambda_4'\,, \qquad \text{(C.64)}
$$

$$
\begin{aligned}
(T\Lambda_2)(T\Lambda_2) &= T^2\,(\Lambda_4 - \Psi_0)\,(\Lambda_4 - \Psi_0) = T^2\,(\Lambda_4\Lambda_4 - \Lambda_4\Psi_0 - \Psi_0\Lambda_4 + \Psi_0\Psi_0) \\
&= T^2\left(\frac{1}{T}\Lambda_4 - \frac{2}{T}\Psi_0 + \frac{1}{T}\Psi_0\right) = T\,(\Lambda_4 - \Psi_0) = T\Lambda_2 = T\Lambda_2'\,, \qquad \text{(C.65)}
\end{aligned}
$$

$$
(T\Psi_R)(T\Psi_R) = \{\bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z]\}\{\bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z]\} = \bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z] = T\Psi_R = T\Psi_R'\,, \qquad \text{(C.66)}
$$

$$(T\Lambda_R)(T\Lambda_R) = \bar{\mathrm{M}}[Z]\,\bar{\mathrm{M}}[Z] = \bar{\mathrm{M}}[Z] = T\Lambda_R = T\Lambda_R'\,. \tag{C.67}$$

□

Proof of Theorem 4.2 Using Lemma C.1, we first note the following identities :

$$B_1Y = (Y'M_1Y)^{-1}Y'M_1Y = I_G = (Y'N_1Y)^{-1}Y'N_1Y = B_2Y\,, \tag{C.68}$$

$$\bar{\mathrm{M}}[M_1Y]\,M_1\,Y = \bar{\mathrm{M}}[N_1Y]\,N_1\,Y = 0\,, \quad B_1X_1 = B_2X_1 = 0\,, \quad N_1\,X_1 = M_1\,X_1 = 0\,, \tag{C.69}$$

$$N_2M_1Y = (I_T - M_1YA_2)M_1Y = (M_1 - M_1YA_2)Y = M_1(Y - YA_2Y) = 0\,, \quad N_2M_1X_1 = 0\,, \tag{C.70}$$

$$\bar{\mathrm{M}}[\bar{Y}]Y = \bar{\mathrm{M}}[Z]Y = 0\,,\ \bar{\mathrm{M}}[\bar{Y}]X_1 = \bar{\mathrm{M}}[Z]X_1 = 0\,,\ \bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\bar{\mathrm{M}}[\bar{Y}] = \bar{\mathrm{M}}[\bar{Y}]\,\bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\bar{\mathrm{M}}[\bar{Y}]\,. \tag{C.71}$$

Then

$$C_1\,y = (B_2 - B_1)(Y\beta + X_1\gamma + u) = C_1u\,, \tag{C.72}$$

$$y'\,\Psi_0\,y = y'\,C_1'\hat{\Delta}^{-1}C_1\,y = u'\,C_1'\hat{\Delta}^{-1}C_1\,u = u'\,\Psi_0\,u\,, \tag{C.73}$$

$$y'\,\Lambda_1\,y = \frac{1}{T}y'\,N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1\,y = \frac{1}{T}u'\,N_1\,\bar{\mathrm{M}}[N_1Y]\,N_1\,u = u'\,\Lambda_1\,u\,, \tag{C.74}$$

$$y'\,\Lambda_2\,y = \frac{1}{T}y'\,M_1\,(\bar{\mathrm{M}}[M_1Y] - \Psi_0)\,M_1\,y = \frac{1}{T}u'\,M_1\,(\bar{\mathrm{M}}[M_1Y] - \Psi_0)\,M_1\,u = u'\,\Lambda_2\,u\,, \tag{C.75}$$

$$y'\,\Lambda_3\,y = \frac{1}{T}y'\,M_1\,N_2'N_2M_1\,y = \frac{1}{T}u'\,M_1\,N_2'N_2M_1\,u\,, \tag{C.76}$$

$$y'\,\Lambda_4\,y = \frac{1}{T}y'\,\bar{\mathrm{M}}[\bar{Y}]\,y = \frac{1}{T}u'\,\bar{\mathrm{M}}[\bar{Y}]\,u = u'\,\Lambda_4\,u\,, \tag{C.77}$$

$$\begin{aligned} y'\,\Psi_R\,y &= \frac{1}{T}y'\,\bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\,\bar{\mathrm{M}}[\bar{Y}]\,y = \frac{1}{T}y'\,\bar{\mathrm{M}}[\bar{Y}]\,\bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\,\bar{\mathrm{M}}[\bar{Y}]\,y \\ &= \frac{1}{T}u'\,\bar{\mathrm{M}}[\bar{Y}]\,\bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\,\bar{\mathrm{M}}[\bar{Y}]\,u = \frac{1}{T}u'\,\bar{\mathrm{P}}\big[\bar{\mathrm{M}}[\bar{Y}]X_2\big]\,\bar{\mathrm{M}}[\bar{Y}]\,u = u'\,\Psi_R\,u\,, \end{aligned} \tag{C.78}$$

$$\hat{\sigma}_R^2 = \frac{1}{T}y'\,\bar{\mathrm{M}}[Z]y = \frac{1}{T}u'\,\bar{\mathrm{M}}[Z]\,u\,. \tag{C.79}$$

Further, when $a = 0$, we have $u = \sigma_1(\bar{X})\,\varepsilon$, and the expressions in (4.7) - (4.8) follow from (4.1) - (4.3) in Proposition 4.1 once u is replaced by $\sigma_1(\bar{X})\,\varepsilon$ in (C.72) - (C.79). $\sigma_1(\bar{X})$ disappears because it can be factorized in both the numerator and the denominator of each statistic. □

Proof of Proposition 5.1 We must study how the statistics defined in (3.1) - (3.3) change when y and Y are replaced by $y^* = yR_{11} + YR_{21}$ and $Y^* = YR_{22}$. This can be done by looking at the way the relevant variables in (3.4) - (3.14) change. We first note that

$$\hat{\Omega}_{IV}^{*}=\frac{1}{T}Y^{*\prime}N_{1}Y^{*}=(YR_{22})^{\prime}N_{1}(YR_{22})=R_{22}^{\prime}\hat{\Omega}_{IV}R_{22}\,,\quad \hat{\Omega}_{LS}^{*}=\frac{1}{T}Y^{*\prime}M_{1}Y^{*}=R_{22}^{\prime}\hat{\Omega}_{LS}R_{22}\,, \tag{C.80}$$

hence

$$\hat{\Delta}^{*}=(\hat{\Omega}_{IV}^{*})^{-1}-(\hat{\Omega}_{LS}^{*})^{-1}=R_{22}^{-1}(\hat{\Omega}_{IV}^{-1}-\hat{\Omega}_{LS}^{-1})(R_{22}^{-1})^{\prime}=R_{22}^{-1}\hat{\Delta}(R_{22}^{-1})^{\prime}\,. \tag{C.81}$$

Using Lemma C.1, we also get :

$$\begin{aligned} B_{1}^{*} &= (Y^{*\prime}M_{1}Y^{*})^{-1}Y^{*\prime}M_{1}=[(YR_{22})^{\prime}M_{1}(YR_{22})]^{-1}(YR_{22})^{\prime}M_{1}=R_{22}^{-1}(Y^{\prime}M_{1}Y)^{-1}Y^{\prime}M_{1} \\ &= R_{22}^{-1}B_{1}\,, \end{aligned} \tag{C.82}$$

$$B_{2}^{*}=(Y^{*\prime}N_{1}Y^{*})^{-1}Y^{*\prime}N_{1}=R_{22}^{-1}(Y^{\prime}N_{1}Y)^{-1}Y^{\prime}N_{1}=R_{22}^{-1}B_{2}\,, \tag{C.83}$$

$$C_{1}^{*}=B_{2}^{*}-B_{1}^{*}=R_{22}^{-1}C_{1}\,,\quad C_{1}^{*}Y=R_{22}^{-1}C_{1}Y=0\,, \tag{C.84}$$

$$\hat{\beta}^{*}=B_{1}^{*}y^{*}=R_{22}^{-1}B_{1}(yR_{11}+YR_{21})=R_{11}R_{22}^{-1}\hat{\beta}+R_{22}^{-1}R_{21}\,, \tag{C.85}$$

$$\tilde{\beta}^{*}=B_{2}^{*}y^{*}=R_{11}R_{22}^{-1}\tilde{\beta}+R_{22}^{-1}R_{21}\,, \tag{C.86}$$

$$\tilde{\beta}^{*}-\hat{\beta}^{*}=C_{1}^{*}y^{*}=R_{11}R_{22}^{-1}(\tilde{\beta}-\hat{\beta})\,, \tag{C.87}$$

$$\begin{aligned} \hat{u}^{*} &= M_{1}(y^{*}-Y^{*}\hat{\beta}^{*})=M_{1}\big(yR_{11}+YR_{21}-YR_{22}(R_{11}R_{22}^{-1}\hat{\beta}+R_{22}^{-1}R_{21})\big) \\ &= R_{11}M_{1}(y-Y\hat{\beta})=R_{11}\,\hat{u}\,, \end{aligned} \tag{C.88}$$

$$\tilde{u}^{*}=M_{1}(y^{*}-Y^{*}\tilde{\beta}^{*})=M_{1}\big(yR_{11}+YR_{21}-YR_{22}(R_{11}R_{22}^{-1}\tilde{\beta}+R_{22}^{-1}R_{21})\big)=R_{11}\,\tilde{u}\,, \tag{C.89}$$

hence, since $N_{1}X_{1}=0$,

$$\hat{\sigma}^{*2}=\frac{1}{T}\hat{u}^{*\prime}\hat{u}^{*}=R_{11}^{2}\,\hat{\sigma}^{2}\,,\quad \tilde{\sigma}^{*2}=\frac{1}{T}\tilde{u}^{*\prime}\tilde{u}^{*}=R_{11}^{2}\,\tilde{\sigma}^{2}\,, \tag{C.90}$$

$$\begin{aligned} \tilde{\sigma}_{1}^{*2} &= \frac{1}{T}(y^{*}-Y^{*}\tilde{\beta}^{*})^{\prime}N_{1}(y^{*}-Y^{*}\tilde{\beta}^{*})=\frac{1}{T}(y^{*}-Y^{*}\tilde{\beta}^{*}-X_{1}\tilde{\gamma}^{*})^{\prime}N_{1}(y^{*}-Y^{*}\tilde{\beta}^{*}-X_{1}\tilde{\gamma}^{*}) \\ &= \frac{1}{T}\tilde{u}^{*\prime}N_{1}\,\tilde{u}^{*}=R_{11}^{2}\frac{1}{T}\tilde{u}^{\prime}N_{1}\,\tilde{u}=R_{11}^{2}\tilde{\sigma}_{1}^{2}\,, \end{aligned} \tag{C.91}$$

$$\begin{aligned} \tilde{\sigma}_{2}^{*2} &= \hat{\sigma}^{*2}-(\tilde{\beta}^{*}-\hat{\beta}^{*})^{\prime}(\hat{\Delta}^{*})^{-1}(\tilde{\beta}^{*}-\hat{\beta}^{*}) \\ &= R_{11}^{2}\,\hat{\sigma}^{2}-(\tilde{\beta}-\hat{\beta})^{\prime}(R_{11}R_{22}^{-1})^{\prime}R_{22}^{\prime}\hat{\Delta}^{-1}R_{22}(R_{11}R_{22}^{-1})(\tilde{\beta}-\hat{\beta}) \\ &= R_{11}^{2}[\hat{\sigma}^{2}-(\tilde{\beta}-\hat{\beta})^{\prime}\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta})]=R_{11}^{2}\,\tilde{\sigma}_{2}^{2}\,, \end{aligned} \tag{C.92}$$

$$\begin{aligned} \tilde{\Sigma}_{i}^{*} &= \tilde{\sigma}_{i}^{*2}\hat{\Delta}^{*}=(R_{11}^{2}\tilde{\sigma}_{i}^{2})R_{22}^{-1}\hat{\Delta}(R_{22}^{-1})^{\prime}=R_{11}^{2}\,R_{22}^{-1}(\tilde{\sigma}_{i}^{2}\hat{\Delta})(R_{22}^{-1})^{\prime} \\ &= R_{11}^{2}\,R_{22}^{-1}\,\tilde{\Sigma}_{i}\,(R_{22}^{-1})^{\prime},\quad i=1,2,3,4\,, \end{aligned} \tag{C.93}$$

$$\hat{\Sigma}_j^* = R_{11}^2 R_{22}^{-1} \hat{\Sigma}_j (R_{22}^{-1})', \quad j = 1, 2, 3. \tag{C.94}$$

It follows that the $\mathscr{T}_i$ and $\mathscr{H}_j$ exogeneity test statistics based on the transformed data are identical to those based on the original data :

$$\begin{aligned} \mathscr{T}_i^* &= \kappa_i(\tilde{\beta}^* - \hat{\beta}^*)'(\tilde{\Sigma}_i^*)^{-1}(\tilde{\beta}^* - \hat{\beta}^*) \\ &= (\tilde{\beta} - \hat{\beta})'(R_{11}R_{22}^{-1})'[R_{11}^2 R_{22}^{-1} \tilde{\Sigma}_i (R_{22}^{-1})']^{-1}(R_{11}R_{22}^{-1})(\tilde{\beta} - \hat{\beta}) \\ &= \kappa_i(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_i^{-1}(\tilde{\beta} - \hat{\beta}) = \mathscr{T}_i, \; i = 1, 2, 3, 4, \end{aligned} \tag{C.95}$$

$$\begin{aligned} \mathscr{H}_j^* &= T(\tilde{\beta}^* - \hat{\beta}^*)'(\hat{\Sigma}_j^*)^{-1}(\tilde{\beta}^* - \hat{\beta}^*) \\ &= T(\tilde{\beta} - \hat{\beta})'(R_{11}R_{22}^{-1})'[R_{11}^2 R_{22}^{-1} \hat{\Sigma}_j (R_{22}^{-1})']^{-1}(R_{11}R_{22}^{-1})(\tilde{\beta} - \hat{\beta}) = \mathscr{H}_j, \; j = 1, 2, 3. \end{aligned} \tag{C.96}$$

Finally, the invariance of the statistic R is obtained by observing that

$$y^{*\prime}\bar{\mathrm{M}}[Z^*]y^* = R_{11}^2 y'\bar{\mathrm{M}}[Z]y, \quad y^{*\prime}\bar{\mathrm{M}}[\bar{Y}^*]y^* = R_{11}^2 y'\bar{\mathrm{M}}[\bar{Y}]y, \tag{C.97}$$

where $Z^* = [Y^*, X_1, X_2]$ and $\bar{Y}^* = [Y^*, X_1]$, so R_{11}^2 cancels out in R □

PROOF OF THEOREM 6.1 Since $u = Va + \sigma_1(\bar{X})\,\varepsilon$, we can use the identities (C.72) - (C.79) and replace y by $Va + \sigma_1(\bar{X})\,\varepsilon$ in (4.1) - (4.1). The expressions (6.2) - (6.4) then follow through division of the numerator and denominator of each statistic by $\sigma_1(\bar{X})$. □

PROOF OF THEOREM 6.2 This result follows by applying the invariance property of Proposition 5.1 with R defined as in (5.2). y is then replaced by $y^* = X_1\gamma + [V - g(X_1, X_2, X_3, V, \bar{\Pi})]a + e$ [see (5.5)], and the identities (C.72) - (C.79) hold with u replaced by

$$u_* = [V - g(X_1, X_2, X_3, V, \bar{\Pi})]a + e\,. \tag{C.98}$$

Further, in view of (C.5) and (4.4) - (3.14), each one of the matrices Ψ_0, $\Lambda_1, \ldots, \Lambda_4$, Ψ_1, Ψ_R and Λ_R remains the same if it is pre- and postmultiplied by M_1, *i.e.*

$$\Psi_0 = M_1\Psi_0 M_1\,, \quad \Lambda_l = M_1\Lambda_i M_1,\; i = 1, 2, 3, 4, \tag{C.99}$$

$$\Psi_1 = M_1\Psi_1 M_1\,, \quad \Psi_R = M_1\Psi_R M_1\,, \quad \Lambda_R = M_1\Lambda_R M_1\,, \tag{C.100}$$

so u_* can in turn be replaced by

$$M_1 u_* = -M_1[V - g(X_1, X_2, X_3, V, \bar{\Pi})]a + M_1 e \tag{C.101}$$

in (C.72) - (C.79). Upon division of the numerator and denominator of each statistic by $\sigma_1(\bar{X})$, we get the expressions (6.6) - (6.8). □

PROOF OF THEOREM 6.3 The result follows from well known properties of the normal and chi-square distributions: if $x \sim N_n[\mu, I_n]$ and A is a fixed idempotent $n \times n$ matrix of rank r, then $x'Ax \sim \chi^2[r; \mu' A \mu]$. Conditional on $\bar{X}$ and V, Ψ_0 is fixed, and

$$y_*^{\perp}(\bar{a}) = \bar{\mu}_{y*}^{\perp}(\bar{a}) + M_1 \varepsilon = M_1\{[V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a} + \varepsilon\} = M_1(\mu + \varepsilon) \tag{C.102}$$

where $\mu = [V - g(X_1, X_2, X_3, V, \bar{\Pi})]\bar{a}$ is fixed and $\varepsilon \sim N_n[\mu, I_n]$. By Lemmas C.1 and C.2, $T\Psi_0$, $T\Lambda_1$, $T\Lambda_2$, $T\Lambda_4$, $T\Psi_R$ and $T\Lambda_R$ are symmetric idempotent, and each of these matrices remain invariant through by pre- and post-multiplication by M_1 $[M_1 \Psi_0 M_1 = \Psi_0$, etc.]. Thus

$$\begin{aligned} S_T[y_*^{\perp}(\bar{a}), \Psi_0] &= T\, y_*^{\perp}(\bar{a})' \Psi_0 y_*^{\perp}(\bar{a}) = (\mu + \varepsilon)' M_1 (T\Psi_0) M_1 (\mu + \varepsilon) && \text{(C.103)} \\ &= (\mu + \varepsilon)'(T\Psi_0)(\mu + \varepsilon) \sim \chi^2[\text{rank}(T\Psi_0); \mu'(T\Psi_0)\mu] && \text{(C.104)} \end{aligned}$$

where

$$\text{rank}(T\Psi_0) = \text{tr}(T\Psi_0) = \text{tr}(T C_1' \hat{\Delta}^{-1} C_1) = \text{tr}(T \hat{\Delta}^{-1} C_1 C_1') = \text{tr}(T \hat{\Delta}^{-1} T^{-1} \hat{\Delta}) = G, \tag{C.105}$$

$$\mu'(T\Psi_0)\mu = \mu' M_1 (T\Psi_0) M_1 \mu = \bar{\mu}_{y*}^{\perp}(\bar{a})'(T\Psi_0)\bar{\mu}_{y*}^{\perp}(\bar{a}) = S_T[\bar{\mu}_{y*}^{\perp}(\bar{a}), \Psi_0] = \delta(\bar{a}, \Psi_0). \tag{C.106}$$

The proofs for the other quadratic forms are similar, with the following degrees of freedom vary:

$$\begin{aligned} \text{rank}(T\Lambda_1) &= \text{tr}\{N_1 \bar{\mathrm{M}}[N_1 Y] N_1\} = \text{tr}\{N_1\} - \text{tr}\{\bar{\mathrm{P}}[N_1 Y]\} = \text{tr}\{M_1 - M\} - \text{tr}\{N_1 Y (Y' N_1 Y)^{-1} Y' N_1\} \\ &= (T - k_1) - (T - k_1 - k_2) - \text{tr}\{(Y' N_1 Y)^{-1} Y' N_1 Y\} = k_2 - G, && \text{(C.107)} \end{aligned}$$

$$\begin{aligned} \text{rank}(T\Lambda_2) &= \text{tr}\{T M_1 \left(T^{-1}\bar{\mathrm{M}}[M_1 Y] - \Psi_0\right) M_1\} = \text{tr}\{M_1 \bar{\mathrm{M}}[M_1 Y] M_1\} - \text{tr}\{T\Psi_0\} \\ &= \text{tr}\{M_1\} - \text{tr}\{\bar{\mathrm{P}}[M_1 Y]\} - \text{tr}\{T\Psi_0\} = T - k_1 - 2G, && \text{(C.108)} \end{aligned}$$

$$\text{rank}(T\Lambda_4) = \text{tr}\{M_1 \bar{\mathrm{M}}[M_1 Y] M_1\} = \text{tr}\{M_1\} - \text{tr}\{\bar{\mathrm{P}}[M_1 Y]\} = T - k_1 - G, \tag{C.109}$$

$$\text{rank}(T\Psi_R) = \text{tr}\{\bar{\mathrm{M}}[\bar{Y}] - \bar{\mathrm{M}}[Z]\} = (T - k_1 - G) - (T - k_1 - G - k_2) = k_2, \tag{C.110}$$

$$\text{rank}(T\Lambda_R) = \text{tr}(T\Lambda_R) = \text{tr}\{\bar{\mathrm{M}}[Z]\} = T - G - k_1 - k_2. \tag{C.111}$$

The independence properties follow from the orthogonalities given in (C.49) and the normality assumption. □

PROOF OF COROLLARY 6.4 These results directly from Theorem 6.3 and the definition of the doubly noncentral F-distribution. □

References

Ahn, S. C. (1997), 'Orthogonality tests in linear models', *Oxford Bulletin of Economics and Statistics* **59**, 83–186.

Andrews, D. W. K. and Stock, J. H. (2007), Inference with weak instruments, *in* R. Blundell, W. Newey and T. Persson, eds, 'Advances in Economics and Econometrics Theory and Applications, Ninth World Congress', Vol. 3, Cambridge University Press, Cambridge, U.K., chapter 6.

Angrist, J. D., Imbens, G. W. and Krueger, A. B. (1999), 'Jackknife instrumental variables estimates', *Journal of Applied Econometrics* **14**, 57–67.

Angrist, J. D. and Krueger, A. B. (1991), 'Does compulsory school attendance affect schooling and earning?', *Quarterly Journal of Economics* **CVI**, 979–1014.

Angrist, J. D. and Krueger, A. B. (1995), 'Split-sample instrumental variables estimates of the return to schooling', *Journal of Business and Economic Statistics* **13**, 225–235.

Angrist, J. D. and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, Princeton, New Jersey.

Bahadur, R. R. and Savage, L. J. (1956), 'The nonexistence of certain statistical procedures in nonparametric problems', *Annals of Mathematical Statistics* **27**(4), 1115–1122.

Baum, C., Schaffer, M. and Stillman, S. (2003), 'Instrumental variables and GMM: Estimation and testing', *Stata Journal* **3**(1), 1–30.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2005), Exact multivariate tests of asset pricing models with stable asymmetric distributions, *in* M. Breton and H. Ben Ameur, eds, 'Numerical Methods in Finance', Kluwer/Springer-Verlag, New York, chapter 9, pp. 173–191.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2007), 'Multivariate tests of mean-variance efficiency with possibly non-Gaussian errors: An exact simulation-based approach', *Journal of Business & Economic Statistics* **25**(4), 398–410.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2009), 'Finite-sample multivariate tests of asset pricing models with coskewness', *Computational Statistics and Data Analysis* **53**, 2008–2021.

Beaulieu, M.-C., Dufour, J.-M. and Khalaf, L. (2013), 'Identification-robust estimation and testing of the zero-beta CAPM', *Review of Economic Studies,* **83**(3), 892–924.

Blundell, R., R. and Horowitz, J. L. (2007), 'A non-parametric test of exogeneity', *Review of Economic Studies* **74**, 1035–1058.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**, 443–450.

Caetano, C. (2015), 'A test of exogeneity without instrumental variables in models with bunching', *Econometrica* **83**(4), 1581–1600.

Chmelarova, V. and Hill, R. C. (2010), 'The Hausman pretest estimator', *Economics Letters* **108**, 96–99.

Davidson, R., Godfrey, L. and Mackinnon, J. G. (1985), 'A simplified version of the differencing test', *International Economic Review* **26**(3), 639–647.

Davidson, R. and Mackinnon, J. G. (1985), 'The interpretation of test statistics', *Canadian Journal of Economics* **18**(1), 38–57.

Davidson, R. and Mackinnon, J. G. (1989), 'Testing for consistency using artificial regressions', *Econometric Theory* **5**(3), 363–384.

Davidson, R. and Mackinnon, J. G. (1990), 'Specification tests based on artificial regressions', *Journal of the American Statistical Association* **85**(409), 220–227.

Davidson, R. and Mackinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York, New York.

Doko Tchatoka, F. (2015*a*), 'On bootstrap validity for specification tests with weak instruments', *The Econometrics Journal* **31**(6), 137–146.

Doko Tchatoka, F. (2015*b*), 'Subset hypotheses testing and instrument exclusion in the linear IV regression', *Econometric Theory* **18**(1), 1192–1228.

Doko Tchatoka, F. and Dufour, J.-M. (2014), 'Identification-robust inference for endogeneity parameters in linear structural models', *The Econometrics Journal* **17**, 165–187.

Doko Tchatoka, F. and Dufour, J.-M. (2016), Exogeneity tests and weak identification in IV regressions: Asymptotic theory and point estimation, Technical report, Department of Economics, McGill University, Montréal, Canada.

Dufour, J.-M. (1979), Methods for Specification Errors Analysis with Macroeconomic Applications, PhD thesis, University of Chicago. 257 + XIV pages.

Dufour, J.-M. (1987), Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations, *in* I. B. MacNeill and 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', *Journal of Econometrics* **133**(2), 443–477.

Dufour, J.-M., Farhat, A., Gardiol, L. and Khalaf, L. (1998), 'Simulation-based finite sample normality tests in linear regressions', *The Econometrics Journal* **1**, 154–173.

Dufour, J.-M. and Hsiao, C. (2008), Identification, *in* L. E. Blume and S. N. Durlauf, eds, 'The New Palgrave Dictionary of Economics', second edn, Palgrave Macmillan, Basingstoke, Hampshire, England.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2003), 'Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models', *Oxford Bulletin of Economics and Statistics* **65**, 891–906.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2013), 'Identification-robustanalysis of DSGE and structural macroeconomic models', *Journal of Monetary Economics* **60**, 340–350.

Dufour, J.-M. and 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.

Engle, R. F. (1982), 'A general approach to Lagrange multiplier diagnostics', *Journal of Econometrics* **20**, 83–104.

Engle, R. F., Hendry, D. F. and Richard, J.-F. (1982), 'Exogeneity', *Econometrica* **51**, 277–304.

Farebrother, R. W. (1976), 'A remark on the Wu test', *Econometrica* **44**, 475–477.

Frankel, J. A. and Romer, D. (1999), 'Does trade cause growth?', *American Economic Review* **89**(3), 379–399.

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. and Moon, H. R. (2010), 'The Hausman test and weak instruments', *Journal of Econometrics* **160**, 289–299.

Hahn, J. and Hausman, J. (2002), 'A new specification test for the validity of instrumental variables', *Econometrica* **70**, 163–189.

Harrison, A. (1996), 'Openness and growth: a time-series, cross-country analysis for developing countries', *Journal of Development Economics* **48**, 419–447.

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.

Hausman, J. and Taylor, W. E. (1981), 'A generalized specification test', *Economics Letters* **8**, 239–245.

Holly, A. (1982), 'A remark on Hausman's test', *Econometrica* **50**, 749–759.

Holly, A. (1983*a*), 'Tests d'exogénéité dans un modèle à equations simultanées: Énoncé de résultats théoriques en information limitée et illustrations à des tests de dépendance de la politique monétaire en régime de changes fixes', *Cahiers du Séminaire d'Économétrie* **25**, 49–69.

Holly, A. (1983*b*), 'Une présentation unifiée des tests d'exogénéity dans les modèles à equations simultanées', *Annales de l'INSEE* **50**, 3–24.

Holly, A. and Monfort, A. (1983), 'Some useful equivalence properties of Hausman's test', *Economics Letters* **20**, 39–43.

Hwang, H.-S. (1980), 'Test of independence between a subset of stochastic regressors and disturbances', *International Economic Review* **21**, 749–760.

Hwang, H.-S. (1985), 'The equivalence of Hausman and Lagrange multiplier tests of independence between disturbance and a subset of stochastic regressors', *Economics Letters* **17**, 83–86.

Irwin, A.-D. and Tervio, M. (2002), 'Does trade raise income? Evidence from Twentieth Century', *Journal of International Economics* **58**, 1–18.

Jeong, J. and Yoon, B. H. (2010), 'The effect of pseudo-exogenous instrumental variables on Hausman test', *Communications in Statistics: Simulation and Computation* **39**, 315–321.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995), *Continuous Univariate Distributions, Volume 2*, second edn, John Wiley & Sons, New York.

Kabaila, P., Mainzer, R. and Farchione, D. (2015), 'The impact of a Hausman pretest, applied to panel data, on the coverage probability of confidence intervals', *Economics Letters* **131**, 12–15.

Kariya, T. and Hodoshima, H. (1980), 'Finite-sample properties of the tests for independence in structural systems and LRT', *The Quarterly Journal of Economics* **31**, 45–56.

Kiviet, J. F. (2013), 'Identification and inference in a simultaneous equation under alternative information sets and sampling schemes', *The Econometrics Journal* **16**, S24–S59.

Kiviet, J. F. and Niemczyk, J. (2006), On the limiting and empirical distribution of IV estimators when some of the instruments are invalid, Technical report, Department of Quantitative Economics, University of Amsterdam, Amsterdam, The Netherlands.

Kiviet, J. F. and Niemczyk, J. (2007), 'The asymptotic and finite-sample distributions of OLS and simple IV in simultaneous equations', *Computational Statistics and Data Analysis* **51**, 3296–3318.

Kiviet, J. F. and Pleus, M. (2012), The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation, Technical report, Amsterdam School of Economics, Amsterdam, The Netherlands.

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

Kleibergen, F. and Mavroeidis, S. (2009), 'Rejoinder', *Journal of Business and Economic Statistics* **27**, 331–339.

Lee, Y. and Okui, R. (2012), 'Hahn-Hausman test as a specification test', *Journal of Econometrics* **167**, 133–139.

Lochner, L. and Moretti, E. (2015), 'Estimating and testing models with many treatment levels and limited instruments', *Review of Economics and Statistics* **97**(2), 387–397.

Mankiw, N. G., Romer, D. and Weil, D. N. (1992), 'A contribution to the empirics of economic growth', *The Quarterly Journal of Economics* **107**(2), 407–437.

Meepagala, G. (1992), 'On the finite sample performance of exogeneity tests of Revankar, Revankar and Hartley and Wu-Hausman', *Econometric Reviews* **11**, 337–353.

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

Nakamura, A. and Nakamura, M. (1981), 'On the relationships among several specification error tests presented by Durbin, Wu and Hausman', *Econometrica* **49**, 1583–1588.

Nakamura, A. and Nakamura, M. (1985), 'On the performance of tests by Wu and by Hausman for detecting the ordinary least squares bias problem', *Journal of Econometrics* **29**, 213–227.

Newey, W. K. (1985*a*), 'Generalized method of moments specification testing', *Journal of Econometrics* **29**, 229–256.

Newey, W. K. (1985*b*), 'Maximum likelihood specification testing and conditional moment tests', *Econometrica* **53**(5), 1047–1070.

Pearl, J. (2009), *Causality: Models, Reasoning, and Inference*, second edn, Cambridge University Press, Cambridge, U.K.

Poskitt, D. S. and Skeels, C. L. (2012), 'Inference in the presence of weak instruments: A selected survey', *FTEcx* **6**(1), 26–44.

Revankar, N. S. (1978), 'Asymptotic relative efficiency analysis of certain tests in structural systems', *International Economic Review* **19**, 165–179.

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

Reynolds, R. A. (1982), 'Posterior odds for the hypothesis of independence between stochastic regressors and disturbances', *International Economic Review* **23**(2), 479–490.

Rivers, D. and Vuong, Q. (1988), 'Limited information estimators and exogeneity tests for simultaneous probit models', *Journal of Econometrics* **39**(3), 347–366.

Ruud, P. A. (1984), 'Tests of specification in econometrics', *Econometric Reviews* **3**(2), 211–242.

Ruud, P. A. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press, Inc., New York.

Simon, H. A. (1953), Causal ordering and identifiability, *in* W. C. Hood and T. C. Koopmans, eds, 'Studies in Econometric Method', number 14 *in* 'Cowles Commission Monographs', John Wiley & Sons, New York, chapter III, pp. 49–74.

Smith, R. J. (1983), 'On the classical nature of the Wu-Hausman statistics for independence of stochastic regressors and disturbance', *Economics Letters* **11**, 357–364.

Smith, R. J. (1984), 'A note on likelihood ratio tests for the independence between a subset of stochastic regressors and disturbances', *International Economic Review* **25**, 263–269.

Smith, R. J. (1985), 'Wald tests for the independence of stochastic variables and disturbance of a single linear stochastic simultaneous equation', *Economics Letters* **17**, 87–90.

Smith, R. J. (1994), 'Asymptotically optimal tests using limited information and testing for exogeneity', *Econometric Theory* **10**, 53–69.

Smith, R. J. and Pesaran, M. (1990), 'A unified approach to estimation and orthogonality tests in linear single-equation econometric models', *Journal of Econometrics* **44**, 41–66.

Spencer, D. E. and Berk, K. N. (1981), 'A limited-information specification test', *Econometrica* **49**, 1079–1085. Erratum, Econometrica, Vol. 50, No. 4 (Jul., 1982), p. 1087.

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

Stock, J. H., Wright, J. H. and 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.

Thurman, W. (1986), 'Endogeneity testing in a supply and demand framework', *Review of Economics and Statistics* **68**(4), 638–646.

Wong, K.-f. (1996), 'Bootstrapping Hausman's exogeneity test', *Economics Letters* **53**, 139–143.

Wong, K.-f. (1997), 'Effect on inference of pretesting the exogeneity of a regressor', *Economics Letters* **56**, 267–271.

Wooldridge, J. M. (2014), 'Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variable', *Journal of Econometrics* **182**(1), 226–234.

Wooldridge, J. M. (2015), 'Control function methods in applied econometrics', *Journal of Human Resources* **50**, 420–445.

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

Wu, D.-M. (1974), 'Alternative tests of independence between stochastic regressors and disturbances: Finite sample results', *Econometrica* **42**, 529–546.

Wu, D.-M. (1983*a*), 'A remark on a generalized specification test', *Economics Letters* **11**, 365–370.

Wu, D.-M. (1983*b*), 'Tests of causality, predeterminedness and exogeneity', *International Economic Review* **24**(3), 547–558.