UvA ECONOMETRICS

Discussion Paper: 2011/13

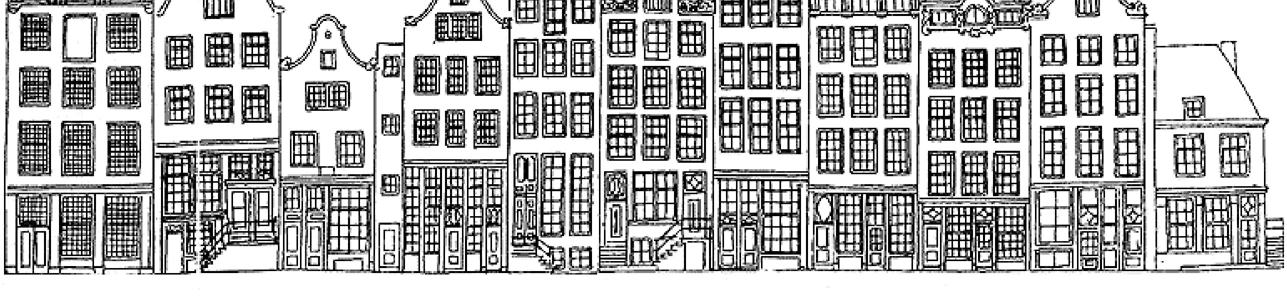
The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation

Jan F. Kiviet and Milan Pleus

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM



The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation

Jan F. Kiviet* and Milan Pleus†

22 March 2016
JEL-code: *C01, C12, C15, C30*

Keywords:
bootstrapping, classification of explanatories, DWH orthogonality tests, test implementation, test performance, simulation design

Abstract

Tests for classification as endogenous or predetermined of arbitrary subsets of regressors are formulated as significance tests in auxiliary IV regressions and their relationships with various more classic test procedures are examined and critically compared with statements in the literature. Then simulation experiments are designed by solving the data generating process parameters from salient econometric features, namely: degree of simultaneity and multicollinearity of regressors, and individual and joint strength of external instrumental variables. Next, for various test implementations, a wide class of relevant cases is scanned for flaws in performance regarding type I and II errors. Substantial size distortions occur, but these can be cured remarkably well through bootstrapping, except when instruments are weak. The power of the subset tests is such that they establish an essential addition to the well-known classic full-set DWH tests in a data based classification of individual explanatory variables. This is also illustrated in an empirical example supplemented with hints for practitioners.

*Emeritus Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, P.O. Box 15867, 1001 NJ Amsterdam, The Netherlands (j.f.kiviet@uva.nl). Major parts of this work were done whilst being Visiting Professor at the Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332.

†Senior Analyst at AHTI (Amsterdam Health & Technology Institute), Paasheuvelweg 25, 1105 BP Amsterdam, The Netherlands (m.pleus@ahti.nl), and formerly Amsterdam School of Economics, University of Amsterdam, The Netherlands. Financial support from the Netherlands Organisation for Scientific Research (NWO) grant "Statistical inference methods regarding effectivity of endogous policy measures" is gratefully acknowledged.

1 Introduction

In this study we examine a range of test procedures for the classification of arbitrary subsets of explanatory variables as either endogenous or predetermined in an adequately specified single structural linear model. Correct classification is highly important because misclassification leads to either inefficient or inconsistent estimation. Hausman's principle, which examines the discrepancy between two alternative estimators, is employed directly and also indirectly. The latter leads to various tests formulated as joint significance tests of additional regressors in auxiliary IV regressions. Their relationships are demonstrated with particular forms of classic tests such as Durbin-Wu-Hausman orthogonality tests, Revankar-Hartley covariance tests and Sargan-Hansen overidentification restriction tests. At various points we indicate misconceptions in the relevant literature. We run simulations over a wide class of relevant cases, to find out which versions have best control over type I error probabilities and to get an idea of the power of these tests. This should help to use these tests effectively in practice when trying to avoid both evils of inconsistency and inefficiency. To that end a simulation approach is developed by which relevant data generating processes (DGPs) are designed by deriving the values for their parameters from chosen salient features of the system, namely: degree of simultaneity of individual explanatory variables, degree of multicollinearity between explanatory variables, and individual and joint strength of employed external instrumental variables. This allows scanning the relevant parameter space for flaws in performance regarding type I and II errors of all implementations of the tests and their bootstrapped versions. We find that testing orthogonality by standard methods is impeded for weakly identified regressors. Like bootstrapped tests require resampling under the null, we find here that testing for orthogonality by auxiliary regressions benefits from estimating variances under the null, as in Lagrange multiplier tests, rather than under the alternative, as in Wald-type tests. However, after proper size correction we find that the Wald-type tests exhibit the best power properties. The best performing procedures are also employed in an illustrative empirical example.

Procedures for testing the orthogonality of all possibly endogenous regressors regarding the error term have been developed by Durbin (1954), Wu (1973), Revankar and Hartley (1973), Revankar (1978) and Hausman (1978). Mutual relationships between these are discussed in Nakamura and Nakamura (1981) and Hausman and Taylor (1981). This test problem has been put into a likelihood framework under normality by Holly (1982) and Smith (1983). Most of the papers just mentioned, and in particular Davidson and MacKinnon (1989, 1990), provide a range of implementations for these tests that can easily be obtained from auxiliary regressions. Although this type of inference problem does address one of the basic fundaments of the econometric analysis of observational data, relatively little evidence on the performance of the available tests in finite samples is available. Monte Carlo studies on the performance of some of the implementations in static linear models can be found in Wu (1974), Meepagala (1992), Chmelarova and Carter Hill (2010), Jeong and Yoon (2010), Hahn et al.(2011) and Doko Tchatoka (2014), whereas such results for linear dynamic models are presented in Kiviet (1985).

The more subtle problem of deriving a test for the orthogonality of subsets of the regressors not involving all of the possibly endogenous regressors has also received substantial attention over the last three decades. Nevertheless, generally accepted rules for best practice on how to approach this problem do not seem available yet, or are confusing as we shall see, and not yet supported by any simulation evidence. Self-evidently, though, the situation where one is convinced of the endogeneity of a few of the regressors, but wants to test some other regressors for orthogonality, is of high practical relevance. If orthogonality is established, this permits to use these regressors as instrumental variables, which (if correct) improves the efficiency and the identification situation, because it makes the analysis less dependent on the availability of external instruments. This is important in particular when available external instruments are weak or of doubtful exogeneity status. Testing the orthogonality of subsets of the possibly endogenous regressors was addressed first by Hwang (1980) and next by Spencer and Berk (1981, 1982), Wu (1983), Smith (1984, 1985), Hwang (1985) and Newey (1985), who all suggest various test procedures, some of them asymptotically or even algebraically equivalent. So do Pesaran and Smith (1990), who also provide theoretical arguments regarding an ordering of the power of the various tests, although they are asymptotically equivalent under the null and under local alternatives. Various of the possible sub-set test implementations are paraphrased in Ruud (1984, 2000), Davidson and MacKinnon (1993) and in Baum et al. (2003), and occasionally their relationships with particular forms of Sargan-Hansen (partial-)overidentification test statistics are examined. As we shall show, a few particular situations still call for further analysis and formal proofs, and sometimes results from the studies mentioned above have to be corrected. As far as we know, there are no published simulation results yet on the actual qualities of tests for the exogeneity for arbitrary subsets of the regressors in finite samples.

In this paper we shall try to elucidate the various forms of available test statistics for the endogeneity of subsets of the regressors, demonstrate their origins and their relationships, and also produce solid Monte Carlo results on their performance in single static linear simultaneous models with IID disturbances. That yet no simulation results are available on sub-set tests may be due to the fact that it is not straightforward how one should design a range of appealing and representative experiments. We believe that in this respect the present study, which closely follows the rules set out in Kiviet (2012), may claim originality. Besides exploiting some invariance properties, we choose the remaining parameter values for the DGP indirectly from the inverse relationships between the DGP parameter values and fundamental orthogonal econometric notions. The latter constitute an insightful base for the relevant nuisance parameter space. The present design can easily be extended to cover cases with a more realistic degree of overidentification and number of jointly dependent regressors. Other obvious extensions would be: to include recently developed tests which are specially built to cope with weak instruments, to consider non Gaussian and non IID disturbances, to examine dynamic models, to include tests for the validity (orthogonality) of instruments which are not included in the regression, etc. Regarding all these aspects the present study just offers an initial reference point.

The structure of the paper is as follows. In Section 2, we first define the model's maintained properties and the hypothesis to be tested. Next, in a series of subsections, various routes to develop test procedures are followed and their resulting test statistics are discussed and compared analytically. Section 3 reviews earlier Monte Carlo designs and results regarding orthogonality tests. In Section 4 we set out our approach to obtain DGP parameter values from chosen basic econometric characteristics. A simulation design is obtained to parametrize a synthetic single linear static regression model including two possibly endogenous regressors with an intercept and involving two external instruments. For this design Section 5 presents simulation results for a selection of practically relevant parametrizations. Section 6 produces similar results for bootstrapped versions of the tests, Section 7 provides an empirical case study and Section 8 concludes.

2 Testing the orthogonality of subsets of explanatory variables

2.1 The model and setting

We consider the single linear simultaneous equation model

$$y = X\beta + u, \tag{2.1}$$

with IID unobserved disturbances $u \sim (0, \sigma^2 I_n)$, K-element unknown coefficient vector β, an $n \times K$ regressor matrix X and $n \times 1$ regressand y. We also have an $n \times L$ matrix Z containing sample observations on identifying instrumental variables, so

$$E(Z'u) = 0,\ rank(Z) = L,\ rank(X) = K \text{ and } rank(Z'X) = K. \tag{2.2}$$

In addition, we make asymptotic regularity assumptions to guarantee asymptotic identification of all elements of β and consistency of its IV (or 2SLS) estimator

$$\hat{\beta} = (X'P_Z X)^{-1} X'P_Z y, \tag{2.3}$$

where $P_Z = Z(Z'Z)^{-1}Z'$. Hence, we assume that

$$\operatorname{plim} n^{-1} Z'Z = \Sigma_{Z'Z} \text{ and } \operatorname{plim} n^{-1} Z'X = \Sigma_{Z'X} \tag{2.4}$$

are finite and have full column rank, whereas $\hat{\beta}$ has limiting normal distribution

$$n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \sigma^2 [\Sigma'_{Z'X} \Sigma^{-1}_{Z'Z} \Sigma_{Z'X}]^{-1}\right). \tag{2.5}$$

The matrices X and Z may have some (but not all) columns in common and can therefore be partitioned as

$$X = (Y, Z_1) \text{ and } Z = (Z_1, Z_2), \tag{2.6}$$

where Z_j has L_j columns for $j = 1, 2$. Because the number of columns in Y is $K - L_1 > 0$ we find from $L = L_1 + L_2 \geq K$ that $L_2 > 0$, but we allow $L_1 \geq 0$, so Z_1 may be void. Throughout this paper the model just defined establishes the maintained (unrestrained) hypothesis, which allows Y to contain endogenous variables. Below we will examine particular restrained versions of the maintained hypothesis and develop tests to verify these further limitations. These are not parametric restraints regarding β but involve orthogonality conditions in addition to the L maintained orthogonality conditions embedded in $E(Z'u) = 0$. All these extra orthogonality conditions concern regressors rather than external instrumental variables. Therefore, we consider a partitioning of Y in K_e and K_o columns

$$Y = (Y_e, Y_o), \tag{2.7}$$

where the variables Y_e are maintained as possibly **e**ndogenous, whereas for the K_o variables Y_o their possible **o**rthogonality will be examined, i.e. whether $E(Y_o'u) = 0$ seems to hold. We define the $n \times (L + K_o)$ matrix

$$Z_r = (Z, Y_o), \tag{2.8}$$

which relates to all the orthogonality conditions in the **r**estrained model. Note that (2.2) implies that Z_r has full column rank, provided $n \geq L + K_o$. Now the null and alternative hypotheses that we will examine can be expressed as

$$\begin{aligned} H^0 &: y = X\beta + u, \quad u \sim (0, \sigma^2 I), \quad E(Z_r'u) = 0, \text{ and} \\ H^1 &: y = X\beta + u, \quad u \sim (0, \sigma^2 I), \quad E(Z'u) = 0, \; E(Y_o'u) \neq 0. \end{aligned} \tag{2.9}$$

Hence, H^0 assumes $E(Y_o'u) = 0$.

Under the extended set of orthogonality conditions $E(Z_r'u) = 0$, i.e. under H^0, the restrained IV estimator is

$$\hat{\beta}_r = (X'P_{Z_r}X)^{-1}X'P_{Z_r}y. \tag{2.10}$$

If H^0 is valid this estimator is consistent and, provided $\operatorname{plim} n^{-1}Z_r'Z_r = \Sigma_{Z_r'Z_r}$ exists and is invertible, its limiting normal distribution has variance $\sigma^2[\Sigma_{Z_r'X}'\Sigma_{Z_r'Z_r}^{-1}\Sigma_{Z_r'X}]^{-1}$, which involves an asymptotic efficiency gain over (2.5). However, under the alternative hypothesis H^1 estimator $\hat{\beta}_r$ is inconsistent.

A test for (2.9) should (as always) have good control over its type I error probability[1] and preferably also have high power, in order to prevent the acceptance of an inconsistent estimator. In practice inference on (2.9) usually establishes just one link in a chain of tests to decide on the adequacy of model specification (2.1) and the maintained instruments Z, see for instance Godfrey and Hutton (1994) and Guggenberger (2010). Many of the firm results obtained below require to make the very strong assumptions embedded in (2.1) and (2.2) and leave it to the practitioner to make a balanced use of them within an actual modelling context.

In the derivations to follow we make use of the following three properties of projection matrices, which for any full column rank matrix A is denoted as $P_A =$

[1] An actual type I error probability much larger than the chosen nominal value would more often than intended lead to using an inefficient estimator. A much lower actual type I error than the nominal level would deprive the test from its power hampering the detection of estimator inconsistency.

$A(A'A)^{-1}A'$: For a full column rank matrix $C = (A, B)$ one has (i) $P_A = P_C P_A = P_A P_C$; (ii) $P_C = P_A + P_{M_A B} = P_{(A, M_A B)}$, where $M_A = I - P_A$; (iii) for $C^* = (A^*, B)$, where $A^* = A - BD$ and D an arbitrary matrix of appropriate dimensions, $P_{C^*} = P_B + P_{M_B A^*} = P_B + P_{M_B A} = P_C$.

2.2 The source of any estimator discrepancy

A test based on the Hausman principle focusses on the discrepancy vector

$$\begin{aligned}
\hat{\beta} - \hat{\beta}_r &= (X'P_Z X)^{-1} X' P_Z y - (X' P_{Z_r} X)^{-1} X' P_{Z_r} y \\
&= (X'P_Z X)^{-1} X' P_Z [I - X(X'P_{Z_r} X)^{-1} X' P_{Z_r}] y \\
&= (X'P_Z X)^{-1} (P_Z X)' \hat{u}_r \\
&= (X'P_Z X)^{-1} (P_Z Y_e, P_Z Y_o, Z_1)' \hat{u}_r,
\end{aligned} \tag{2.11}$$

where $\hat{u}_r = y - X\hat{\beta}_r$ denotes the IV residuals obtained under H^0. Although testing whether the discrepancy between these two coefficient estimators is significantly different from zero is not equivalent to testing H^0, we will show that in fact all existing test procedures employ the outcome of this discrepancy to infer on the (in)validity of H^0. Because $(X'P_Z X)^{-1}$ is non-singular $\hat{\beta} - \hat{\beta}_r$ is close to zero if and only if the $K \times 1$ vector $(P_Z Y_e, P_Z Y_o, Z_1)' \hat{u}_r$ is. So, we will examine now when its three sub-vectors

$$Y_e' P_Z \hat{u}_r, \; Y_o' P_Z \hat{u}_r \text{ and } Z_1' \hat{u}_r \tag{2.12}$$

will jointly be close to zero. Note that due to the identification assumptions both $P_Z Y_e$ and $P_Z Y_o$ will have full column rank so cannot be O.

For the IV residuals $\hat{u}_r$ we have $X' P_{Z_r} \hat{u}_r = 0$, and since $P_{Z_r} X = (P_{Z_r} Y_e, Y_o, Z_1)$, this yields

$$Y_e' P_{Z_r} \hat{u}_r = 0, \; Y_o' \hat{u}_r = 0 \text{ and } Z_1' \hat{u}_r = 0. \tag{2.13}$$

Note that the third vector of (2.12) is always zero according to the third equality from (2.13). Using projection matrix property (ii) and the first equality of (2.13), we find for the first vector of (2.12) that

$$Y_e' P_Z \hat{u}_r = Y_e' (P_{Z_r} - P_{M_Z Y_o}) \hat{u}_r = -Y_e' P_{M_Z Y_o} \hat{u}_r,$$

so

$$Y_e' P_Z \hat{u}_r = -Y_e' M_Z Y_o (Y_o' M_Z Y_o)^{-1} Y_o' M_Z \hat{u}_r. \tag{2.14}$$

This K_e element vector will be close to zero when the K_o element vector $Y_o' M_Z \hat{u}_r$ is. Due to the occurrence of the $K_e \times K_o$ matrix $Y_e' M_Z Y_o$ as a first factor in the right-hand side of (2.14), it seems possible that $Y_e' P_Z \hat{u}_r$ may be close to zero too in cases where $Y_o' M_Z \hat{u}_r \neq 0$; we will return to that possibility below. For the second vector of (2.12) we find, upon using the second equality of (2.13), that

$$Y_o' P_Z \hat{u}_r = -Y_o' M_Z \hat{u}_r. \tag{2.15}$$

Hence, the second vector of (2.12) will be close to zero if and only if the vector $Y_o' M_Z \hat{u}_r$ is close to zero.

From the above it follows that $Y_o'M_Z\hat{u}_r$ being close to zero is both necessary and sufficient for the full discrepancy vector (2.11) to be small. Checking whether $Y_o'M_Z\hat{u}_r$ is close to zero corresponds to examining to what degree the variables M_ZY_o do obey the orthogonality conditions, while using $\hat{u}_r$ as a proxy for u, which is asymptotically valid under the extended set of orthogonality conditions. Note that by focussing on M_ZY_o the tested variables Y_o have been purged from their components spanned by the columns of Z. Since these are maintained to be orthogonal with respect to u, they should better be excluded from the test indeed.

Since the inverse matrix in the right-hand side of (2.11) is positive definite, the probability limits of $\hat{\beta}$ and $\hat{\beta}_r$ will be similar if and only if $\operatorname{plim} n^{-1}Y_o'M_Z\hat{u}_r = 0$. Regarding the power of any discrepancy based test of (2.9) it is now of great interest to examine whether it could happen under H^1 to have $\operatorname{plim} n^{-1}Y_o'M_Z\hat{u}_r = 0$. For that purpose we specify the reduced form equations

$$Y_j = Z\Pi_j + (u\gamma_j' + V_j), \text{ for } j \in \{e, o\}, \tag{2.16}$$

where Π_j is an $L \times K_j$ matrix of reduced form parameters, γ_j is a $K_j \times 1$ vector that parametrizes the simultaneity and V_j establishes the components of the zero mean reduced form disturbances which are uncorrelated with u and of course with Z. After this further parametrization the hypotheses (2.9) can now be expressed as $H^0 : \gamma_o = 0$ and $H^1 : \gamma_o \neq 0$. Let $(L + K_o) \times (L + K_o)$ matrix Ψ be such that $\Psi\Psi' = (Z_r'Z_r)^{-1}$. From

$$\begin{aligned} Y_o'M_Z\hat{u}_r &= -Y_o'P_Z[I_n - X(X'P_{Z_r}X)^{-1}X'P_{Z_r}]u \\ &= -Y_o'P_Z[P_{Z_r} - P_{Z_r}X(X'P_{Z_r}X)^{-1}X'P_{Z_r}]u \\ &= -Y_o'P_ZZ_r\Psi[I_{L+K_o} - P_{\Psi'Z_r'X}]\Psi'Z_r'u \end{aligned} \tag{2.17}$$

it follows that $\operatorname{plim} n^{-1}Y_o'M_Z\hat{u}_r = 0$ if $(L + K_o) \times 1$ vector $\operatorname{plim} n^{-1}Z_r'u = \sigma^2(0' \ \gamma_o')'$ is in the column space spanned by $\operatorname{plim} n^{-1}Z_r'X = \Sigma_{Z_r'X}$. This is obviously the case when $\gamma_o = 0$. However, it cannot occur for $\gamma_o \neq 0$, because $(L+K_o)\times 1$ vector $\Sigma_{Z_r'X}c$, with c a $K \times 1$ vector, has its first $L \geq K$ elements equal to zero only for $c = 0$, due to the identification assumptions. This excludes the existence of a vector $c \neq 0$ yielding $\Sigma_{Z_r'X}c = \sigma^2(0' \ \gamma_o')'$ when $\gamma_o \neq 0$, so under asymptotic identification the discrepancy will be nonzero asymptotically when Y_o contains an endogenous variable.

Cases in which the asymptotic identification assumptions are violated are $\Pi_e = C_e/\sqrt{n}$ and/or $\Pi_o = C_o/\sqrt{n}$, where C_e and C_o are matrices of appropriate dimensions with full column rank and all elements fixed and finite.[2] Examining $\Sigma_{Z_r'X}c$ closer yields

$$\Sigma_{Z_r'X}c = \begin{pmatrix} \Sigma_{Z'Z}\Pi_e \\ \Pi_o'\Sigma_{Z'Z}\Pi_e + \Sigma_{V_o'V_e} + \sigma^2\gamma_o\gamma_e' \end{pmatrix} c_1 + \begin{pmatrix} \Sigma_{Z'Z}\Pi_o \\ \Sigma_{Y_o'Y_o} \end{pmatrix} c_2 + \begin{pmatrix} \Sigma_{Z'Z_1} \\ \Pi_o'\Sigma_{Z'Z_1} \end{pmatrix} c_3, \tag{2.18}$$

where $c = (c_1' \ c_2' \ c_3')'$ and $\Sigma_{V_o'V_e} = \operatorname{plim} n^{-1}V_o'V_e$. If only $\Pi_o = C_o/\sqrt{n}$, so when all the instruments Z are weak and asymptotically irrelevant for the set of regressors Y_o whose

[2]Doko Tchatoka (2014) considers a similar situation for the special case $K_e = 0$ and $K_o = 1$.

orthogonality is tested, we can set $c_1 = 0$ and $c_3 = 0$ and then for $c_2 = \sigma^2 \Sigma^{-1}_{Y_o' Y_o} \gamma_o = \sigma^2(\sigma^2 \gamma_o \gamma_o' + \Sigma_{V_o' V_o})^{-1} \gamma_o \neq 0$ we have $\Sigma_{Z_r' X} c = \sigma^2 (0' \ \gamma_o')' \neq 0$, demonstrating that the test will have no asymptotic power. If only $\Pi_e = C_e/\sqrt{n}$, thus all the instruments Z are weak for Y_e, a solution $c \neq 0$ can be found upon taking $c_2 = 0$, $c_3 = 0$ and $c_1 \neq 0$, provided $\Sigma_{V_o' V_e} + \sigma^2 \gamma_o \gamma_e' \neq O$ or Y_e and Y_o are asymptotically not uncorrelated. Only c_3 has to be set at zero to find a solution when Z is weak for both Y_o and Y_e. From (2.18) it can also be established that when from Z_2 at least $K_e + K_o$ instruments are not weak for Y the discrepancy will always be different from zero asymptotically when $\gamma_o \neq 0$.

Using (2.16) we also find $\text{plim}\, n^{-1} Y_e' M_Z Y_o = \Sigma'_{V_o' V_e} + \sigma^2 \gamma_e \gamma_o'$, which demonstrates that the first vector of (2.12) would for $\gamma_o \neq 0$ tend to zero also when $\gamma_e = 0$ while the reduced form disturbances of Y_e and Y_o are uncorrelated. This indicates the plausible result that a discrepancy based test may loose power when Y_e is unnecessarily treated as endogenous and Y_o is establishing a weak instrument for Y_e after partialing out Z.

2.3 Testing based on the source of any discrepancy

Next we examine the implementation of testing closeness to zero of $Y_o' M_Z \hat{u}_r$ in an auxiliary regression. Consider

$$y = X\beta + P_Z Y_o \zeta + u^*, \tag{2.19}$$

where $u^* = u - P_Z Y_o \zeta$. Its estimation by IV employing the instruments Z_r yields coefficients that can be obtained by applying OLS to the second-stage regression of y on $P_{Z_r} X$ and $P_{Z_r} P_Z Y_o = P_Z Y_o$. For ζ partitioned regression yields

$$\hat{\zeta} = (Y_o' P_Z M_{P_{Z_r} X} P_Z Y_o)^{-1} Y_o' P_Z M_{P_{Z_r} X} y, \tag{2.20}$$

where, using rule (i), $Y_o' P_Z M_{P_{Z_r} X} y = Y_o' P_Z [I - X(X' P_{Z_r} X)^{-1} X' P_{Z_r}] y = Y_o' P_Z \hat{u}_r$. Thus, by testing $\zeta = 0$ in (2.19) we in fact examine whether $Y_o' P_Z \hat{u}_r = -Y_o' M_Z \hat{u}_r$ differs significantly from a zero vector, which is indeed what we aim for.[3]

Alternatively, consider the auxiliary regression

$$y = X\beta + M_Z Y_o \xi + v^*, \tag{2.21}$$

where $v^* = u - M_Z Y_o \xi$. Using the instruments Z_r involves here applying OLS to the second-stage regression of y on $P_{Z_r} X$ and $P_{Z_r} M_Z Y_o = P_{Z_r} Y_o - P_{Z_r} P_Z Y_o = Y_o - P_Z Y_o = M_Z Y_o$. This yields

$$\hat{\xi} = (Y_o' M_Z M_{P_{Z_r} X} M_Z Y_o)^{-1} Y_o' M_Z M_{P_{Z_r} X} y, \tag{2.22}$$

where

$$\begin{aligned} Y_o' M_Z M_{P_{Z_r} X} y &= Y_o' M_Z [I - P_{Z_r} X (X' P_{Z_r} X)^{-1} X' P_{Z_r}] y \\ &= Y_o' [I - X(X' P_{Z_r} X)^{-1} X' P_{Z_r}] y - Y_o' P_Z [I - X(X' P_{Z_r} X)^{-1} X' P_{Z_r}] y \\ &= Y_o' M_Z \hat{u}_r. \end{aligned} \tag{2.23}$$

[3]This procedure provides the explicit solution to the exercise posed in Davidson and MacKinnon (1993, p.242).

Thus, like testing $\zeta = 0$ in (2.19), testing $\xi = 0$ in auxiliary regression (2.21) examines the magnitude of $Y_o' M_Z \hat{u}_r$. The estimator for β resulting from (2.21) is

$$\hat{\beta}_r^* = (X' P_{Z_r} M_{M_Z Y_o} P_{Z_r} X)^{-1} X' P_{Z_r} M_{M_Z Y_o} y.$$

Because $P_{Z_r} M_{M_Z Y_o} = P_{Z_r} - P_{Z_r} P_{M_Z Y_o} = P_{Z_r} - (P_Z + P_{M_Z Y_o}) P_{M_Z Y_o} = P_{Z_r} - P_{M_Z Y_o} = P_Z$, we find $\hat{\beta}_r^* = \hat{\beta}$. Hence, the IV estimator of β exploiting the extended set of instruments in the auxiliary model (2.21) equals the unrestrained IV estimator $\hat{\beta}$. Many text books mention this result for the special case $K_e = 0$.

From the above we find that testing whether included possibly endogenous variables Y_o can actually be used effectively as valid extra instruments, can be done as follows: Add them to Z, so use Z_r as instruments, and add at the same time the regressors $M_Z Y_o$ (the reduced form residuals of the alleged endogenous variables Y_o in the maintained model) to the model, and then test their joint significance. When testing $\xi = 0$ in (2.21) by a Wald-type statistic, and assuming for the moment that σ^2 is known, the test statistic is

$$\sigma^{-2} y' P_{M_{P_{Z_r} X} M_Z Y_o} y = \sigma^{-2} y' (M_A - M_C) y, \tag{2.24}$$

where $A = P_{Z_r} X$, $B = M_Z Y_o$ and $C = (A, B)$. Hence, $y' P_{M_{P_{Z_r} X} M_Z Y_o} y$ is equal to the difference between the OLS residual sums of squares of the restricted (by $\xi = 0$) and the unrestricted second stage regressions (2.21). One easily finds that testing $\zeta = 0$ in (2.19) by a Wald-type test yields in the numerator

$$y' P_{M_{P_{Z_r} X} P_Z Y_o} y = y' (M_A - M_{C^*}) y,$$

with again $A = P_{Z_r} X = (P_{Z_r} Y_e, Y_o, Z_1)$, but $C^* = (A, B^*)$ with $B^* = P_Z Y_o$. Although $C^* \neq C$, both span the same sub-space, so $M_C = M_{C^*}$ and thus the two auxiliary regressions lead to numerically equivalent Wald-type test statistics.

Of course, σ^2 is in fact unknown and should be replaced by an estimator that is consistent under the null. There are various options for this. Two rather obvious choices would be $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$ or $\hat{\sigma}_r^2 = \hat{u}_r'\hat{u}_r/n$, giving rise to two under the null (and also under local alternatives) asymptotically equivalent test statistics, both with $\chi^2(K_o)$ asymptotic null distribution. Further asymptotically equivalent variants can be obtained by employing a degrees of freedom correction in the estimation of σ^2 and/or by dividing the test statistic by K_o and then confronting it with critical values from an F distribution with K_o and $n - l$ degrees of freedom with l some finite number, possibly $K + K_o$.

Testing the orthogonality of Y_o and u, while maintaining the endogeneity of Y_e, by a simple χ^2-form statistic and using as in a Wald-type test the estimate $\hat{\sigma}^2$ (without any degrees of freedom correction) from the unrestrained model, will be indicated by W_o. When using the uncorrected restrained estimator $\hat{\sigma}_r^2$, the statistic will be denoted here as D_o. So we have the two archetype test statistics

$$W_o = y' P_{M_{P_{Z_r} X} M_Z Y_o} y / \hat{\sigma}^2 \text{ and } D_o = y' P_{M_{P_{Z_r} X} M_Z Y_o} y / \hat{\sigma}_r^2. \tag{2.25}$$

Using the restrained σ^2 estimator, as in a Lagrange-multiplier-type test under normality, was already suggested in Durbin (1954, p.27), where $K_e = L_1 = 0$ and $K_o = L_2 = 1$.

Before we discuss further options for estimating σ^2 in general sub-set tests, we shall first focus on the special case $K_e = 0$, where the full set of endogenous regressors is tested. Then $\hat{\sigma}_r^2 = y'M_X y/n = \frac{n-K}{n}s^2$ stems from OLS. Wu (1973) suggested for this case four test statistics, indicated as $T_1, ..., T_4$, where

$$T_4 = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}D_o \text{ and } T_3 = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}W_o. \tag{2.26}$$

On the basis of his simulation results Wu recommended to use the monotonic transformation of T_4 (or D_o)

$$T_2 = \frac{T_4}{1 - \frac{K_o}{n-2K_o-L_1}T_4} = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}\frac{D_o}{1 - D_o/n}. \tag{2.27}$$

He showed that under normality of both structural and reduced form disturbances the null distribution of T_2 is $F(K_o, n - 2K_o - L_1)$ in finite samples.[4] Because $K_e = 0$ implies $M_{P_{Z_r}X} = M_X$ we find from (2.24) that in this case

$$\frac{D_o}{1 - D_o/n} = n\frac{y'P_{M_X M_Z Y_o}y}{y'(M_X - P_{M_X M_Z Y_o})y} = n\frac{y'P_{M_X M_Z Y_o}y}{y'M_{(X\ M_Z Y_o)}y} = \frac{y'P_{M_X M_Z Y_o}y}{\ddot{\sigma}^2}.$$

Hence, from the final expression we see that T_2 estimates σ^2 by $\ddot{\sigma}^2 = y'M_{(X\ M_Z Y_o)}y/n$, which is the OLS residual variance of auxiliary regression (2.21). Like $\hat{\sigma}^2$ and $\hat{\sigma}_r^2$, $\ddot{\sigma}^2$ is consistent under the null, because $\operatorname{plim} n^{-1}Y_o'M_Z\hat{u}_r = 0$ implies, after substituting (2.23) in (2.22), that $\operatorname{plim}\hat{\xi} = 0$.

Pesaran and Smith (1990) show that under the alternative

$$\operatorname{plim}\hat{\sigma}^2 \geq \operatorname{plim}\hat{\sigma}_r^2 \geq \operatorname{plim}\ddot{\sigma}^2$$

and then invoke arguments due to Bahadur to expect that T_2 (which uses $\ddot{\sigma}^2$) has better power than T_4 (which uses $\hat{\sigma}_r^2$), whereas both T_2 and T_4 are expected to outperform T_3 (which uses $\hat{\sigma}^2$). However, they did not verify this experimentally. Moreover, because T_2 is a simple monotonic transformation of T_4 when $K_e = 0$, we also know that after a fully successful size correction both should have equivalent power.

Following the same lines of thought for cases where $K_e > 0$, we expect (after proper size correction) D_o to do better than W_o, but Pesaran and Smith (1990) suggest that an even better result may be expected from formally testing $\xi = 0$ in the auxiliary regression (2.21) while exploiting instruments Z_r. This amounts to the $\chi^2(K_o)$ test statistic T_o, which (omitting its degrees of freedom correction) generalizes Wu's T_2 for cases where $K_e \geq 0$, and is given by

$$T_o = y'P_{M_{P_{Z_r}X}M_Z Y_o}y/\ddot{\sigma}^2 = y'(M_A - M_C)y/\ddot{\sigma}^2, \tag{2.28}$$

[4] Wu's T_1 test for case $K_e = 0$, which under normality has a $F(K_o, L_2 - K_o)$ distribution, has a poor reputation in terms of power and therefore we leave it aside.

with

$$\ddot{\sigma}^2 = (y - X\hat{\beta} - M_Z Y_o \hat{\xi})'(y - X\hat{\beta} - M_Z Y_o \hat{\xi})/n. \tag{2.29}$$

Actually, it seems that Pesaran and Smith (1990, p.49) employ a slightly different estimator for σ^2, namely

$$(y - X\hat{\beta} - M_Z Y_o \hat{\xi}^*)'(y - X\hat{\beta} - M_Z Y_o \hat{\xi}^*)/n \tag{2.30}$$

with

$$\hat{\xi}^* = (Y_o' M_Z Y_o)^{-1} Y_o' M_Z (y - X\hat{\beta}). \tag{2.31}$$

However, because OLS residuals are orthogonal to the regressors we have $Y_o' M_Z (y - X\hat{\beta} - M_Z Y_o \hat{\xi}) = 0$, from which it follows that $\hat{\xi} = \hat{\xi}^*$, so their test is equivalent with T_o.

When $K_e > 0$ the three tests W_o, D_o and T_o are not simple monotonic transformations of each other, so they may have genuinely different size and power properties in finite samples. In particular, we find that for

$$\frac{D_o}{1 - D_o/n} = \frac{y' P_C y - y' P_A y}{(\hat{u}_r' \hat{u}_r - y' P_C y + y' P_A y)/n},$$

the denominator in the right-hand expression differs from $\ddot{\sigma}^2$ (unless $K_e = 0$).[5] Using that $\hat{\xi}$ is given by (2.31) we find from (2.29) that $\ddot{\sigma}^2 = \hat{u}' M_{M_Z Y_o} \hat{u}/n \leq \hat{\sigma}^2$, so

$$0 \leq W_o \leq T_o, \tag{2.32}$$

whereas D_0, with $D_o \geq 0$, can be at either side of W_o and T_o.

2.4 Testing based on the discrepancy as such

Direct application of the Hausman (1978) principle yields the test statistic

$$H_o = (\hat{\beta} - \hat{\beta}_r)'[\hat{\sigma}^2 (X' P_Z X)^{-1} - \hat{\sigma}_r^2 (X' P_{Z_r} X)^{-1}]^- (\hat{\beta} - \hat{\beta}_r), \tag{2.33}$$

which uses a generalized inverse for the matrix in square brackets. When σ^2 were known the matrix in square brackets would certainly be singular though semi-positive definite. Using two different σ^2 estimates might lead to nonsingularity but could yield negative test statistics. As is obvious from the above, (2.33) will not converge to a χ^2_K distribution under H^0, but in our framework to one with K_o degrees of freedom, *cf.* Hausman and Taylor (1981). Some further analysis leads to the following.

Let β have separate components as follows from the decompositions

$$X\beta = Y_e \beta_e + Y_o \beta_o + Z_1 \beta_1 = Y \beta_{eo} + Z_1 \beta_1, \tag{2.34}$$

[5] Therefore, the test statistic (54) suggested in Baum et al. (2003, p.26), although asymptotically equivalent to the tests suggested here, is built on an inappropriate analogy with the $K_e = 0$ case. Moreover, in their formulas (53) and (54) Q^* should be the difference between the residual sums of squares of second-stage regressions, precisely as in (2.25). The line below (54) suggests that Q^* is a difference between squared IV residuals (which would mean that Q^* could be negative) of the (un)restricted auxiliary regressions, although their footnote 25 seems to suggest otherwise.

whereas $(X'P_ZX)^{-1}$ has blocks A_{jk}, $j,k=1,2$, where A_{11} is a $K_{eo}\times K_{eo}$ matrix with $K_{eo}=K_e+K_o$. Then we find from (2.11) and (2.13) that

$$\begin{aligned}\hat\beta-\hat\beta_r &= (X'P_ZX)^{-1}\begin{pmatrix} Y'P_Z\hat u_r \\ 0\end{pmatrix} = \begin{bmatrix} A_{11} \\ A_{21}\end{bmatrix} Y'P_Z\hat u_r, \\ \hat\beta_{eo}-\hat\beta_{eo,r} &= A_{11}Y'P_Z\hat u_r. \end{aligned} \tag{2.35}$$

Hence, the discrepancy vector of the two coefficient estimates of just the regressors in Y, but also those of the full regressor matrix X, are both linear transformations of rank K_{eo} of the vector $Y'P_Z\hat u_r$. Therefore it is obvious that the Hausman-type test statistic (2.33) can also be obtained from

$$H_o=(\hat\beta_{eo}-\hat\beta_{eo,r})'[\hat\sigma^2(Y'P_{M_{Z_1}Z_2}Y)^{-1}-\hat\sigma_r^2(Y'P_{M_{Z_1}(Z_2\ Y_o)}Y)^{-1}]^-(\hat\beta_{eo}-\hat\beta_{eo,r}). \tag{2.36}$$

Both test statistics are algebraically equivalent, because of the unique linear relationship

$$\hat\beta-\hat\beta_r=\begin{bmatrix} I_{K_{eo}} \\ A_{21}A_{11}^{-1}\end{bmatrix}(\hat\beta_{eo}-\hat\beta_{eo,r}). \tag{2.37}$$

Calculating (2.36) instead of (2.33) just mitigates the numerical problems.

One now wonders whether an equivalent Hausman-type test can be calculated on the basis of the discrepancy between the estimated coefficients for just the regressors Y_o. This is not the case, because a relationship of the form $(\hat\beta_{eo}-\hat\beta_{eo,r})=G(\hat\beta_o-\hat\beta_{o,r})$, where G is a $K_{eo}\times K_o$ matrix, cannot be found[6]. However, a matrix G can be found such that $(\hat\beta_{eo}-\hat\beta_{eo,r})=G\hat\xi$, indicating that test H_o can be made equivalent to the three distinct tests of the foregoing subsection, provided similar σ^2 estimates are being used. Using (2.14) and (2.15) in (2.35) we obtain

$$\begin{aligned}\hat\beta_{eo}-\hat\beta_{eo,r} &= A_{11}Y'P_Z\hat u_r \\ &= -A_{11}\begin{bmatrix} Y_e'M_ZY_o(Y_o'M_ZY_o)^{-1} \\ I_{K_o}\end{bmatrix}(Y_o'M_ZM_{P_{Z_r}X}M_ZY_o)\hat\xi, \end{aligned} \tag{2.38}$$

because (2.22) and (2.23) yield $Y_o'M_Z\hat u_r=(Y_o'M_ZM_{P_{Z_r}X}M_ZY_o)\hat\xi$. So, under the null hypothesis particular implementations of W_o, D_o, T_o and H_o are equivalent.[7]

[6]Note that Wu (1983) and Hwang (1985) start off by analyzing a test based on the descripancy $\hat\beta_o-\hat\beta_{o,r}$. Both Wu (1983) and Ruud (1984, p.236) wrongly suggest equivalence of such a test with (2.33) and (2.36).

[7]This generalizes the equivalence result mentioned below (22.27) in Ruud (2000, p.581), which just treats the case $K_e=0$. Note, however, that because Ruud starts off from the full discrepancy vector, the transformation he presents is in fact singular and therefore the inverse function mentioned in his footnote 24 is non-unique (the zero matrix may be replaced with any other matrix of the same dimensions). To obtain a unique inverse transformation, one should start off from the coefficient discrepancy for just the regressors Y, and this is found to be nonsingular for $K_e=0$ only.

2.5 Testing based on covariance of structural and reduced form disturbances

Auxiliary regression (2.21) is used to detect correlation of u and V_o (the reduced form disturbances of Y_o) by examing the covariance of the residuals $\hat{u}_r$ and $M_Z Y_o$. This might perhaps be done in a more direct way by augmenting regression (2.1) by the actual reduced form disturbances, giving

$$y = X\beta + (Y_o - Z\Pi_o)\phi + w^*, \tag{2.39}$$

where $w^* = u - (Y_o - Z\Pi_o)\phi$ with ϕ a $K_o \times 1$ vector. Let $Z\Pi_o = Z_1\Pi_{o1} + Z_2\Pi_{o2}$, then (2.39) can be written as

$$\begin{aligned} y &= Y_e\beta_e + Y_o(\beta_o + \phi) + Z_1(\beta_1 - \Pi_{o1}\phi) - Z_2\Pi_{o2}\phi + w^* \\ &= X\beta^* + Z_2\phi^* + w^* \end{aligned} \tag{2.40}$$

in which we may assume that $E(Z'w^*) = 0$, though $E(Y_e'w^*) \neq 0$. However, testing $\phi^* = 0$, which corresponds to $\phi = 0$ in (2.39), through estimating (2.40) consistently is not an option, unless $K_e = 0$. For $K_e > 0$, which is the case of our primary interest here, (2.40) contains all available instruments as regressors, so we cannot instrument Y_e.

For the case $K_e = 0$ the test of $\phi^* = 0$ yields the test of Revankar and Hartley (1973), which is an exact test under normality. When $K_o = L_2$ (just identification) it specializes to Wu's T_2.[8] When $L_2 > K_o$ (overidentification) Revankar (1978) argues that testing the K_o restrictions $\phi = 0$ by testing the L_2 restrictions $\phi^* = 0$ is inefficient. He then suggests to test $\phi = 0$ by a quadratic form in the difference of the least-squares estimator of $\beta_o + \phi$ in (2.40) and the IV estimator of β_o.[9]

From the above we see that the tests on the covariance of disturbances do not have a straight-forward generalization for the case $K_e > 0$. However, a test that comes close to it replaces the $L - L_1$ columns of Z_2 in (2.40) by a set of $L - K$ regressors Z_2^* which span a subspace of Z_2, such that $(P_Z Y_e, Z_1, Z_2^*)$ spans the same space as Z. Testing these $L - K$ exclusion restrictions yields the familiar Sargan-Hansen test for testing all the so-called overidentification restrictions of model (2.1). It is obvious that this test will have power for alternatives in which Z_2 and u are correlated, possibly because some of the variables in Z_2 are actually omitted regressors. In practical situations this type of test, and also Hausman type tests for the orthogonality of particular instruments not included as regressors in the specification[10], are very useful. However, we do not consider such implementations here, because right from the beginning we have chosen a context in which all instruments Z are assumed to

[8] This is proved as follows: Both tests have regressors X under the null, and under the alternative the full column rank matrices $(X, P_Z Y_o)$ and (X, Z_2) respectively. These matrices span the same space when $X = (Y_o, Z_1)$ and $Z = (Z_1, Z_2)$ have the same number of columns.

[9] Meepagala (1992) produces numerical results indicating that the descripancy based tests have lower power than the Revankar and Hartley (1973) test when instruments are weak and than the Revankar (1978) test when the instruments are strong.

[10] See Hahn et al. (2011) for a study on its behaviour under weak instruments.

be uncorrelated with u. This allows focus on tests serving only the second part of the two-part testing procedure as exposed by Godfrey and Hutton (1994), who also highlight the asymptotic independence of these two parts.

2.6 Testing by an incremental Sargan test

The original test of overidentifying restrictions initiated by Sargan (1958) does not enable to infer directly on the orthogonality of individual instrumental variables, but a so-called incremental Sargan test does. It builds on the maintained hypothesis $E(Z'u) = 0$ and can test the orthogonality of additional potential instrumental variables. Choosing for these the included regressors Y_o yields a test statistic for the hypotheses (2.9) given by

$$S_o = \frac{\hat{u}_r' P_{Z_r} \hat{u}_r}{\hat{\sigma}_r^2} - \frac{\hat{u}' P_Z \hat{u}}{\hat{\sigma}^2}, \tag{2.41}$$

which could be negative. When using for both separate Sargan statistics the same σ^2 estimate, and employing $P_Z \hat{u} = (P_Z - P_{P_Z X})y$, the numerator would be

$$\begin{aligned}\hat{u}_r' P_{Z_r} \hat{u}_r - \hat{u}' P_Z \hat{u} &= y'(P_{Z_r} - P_{P_{Z_r} X} - P_Z + P_{P_Z X})y \\ &= y'(P_{M_Z Y_o} + P_{P_Z X} - P_{P_{Z_r} X})y \\ &= y'(P_{(P_Z X, M_Z Y_o)} - P_{P_{Z_r} X})y,\end{aligned}$$

whereas that of W_o and D_o in (2.24) is given by $y'(P_C - P_A)y$, where $C = (A, B)$ with $A = P_{Z_r} X$ and $B = M_Z Y_o$. Equivalence[11] is proved by using general result (iii) on projection matrices, upon taking $A^* = P_Z X$. Using $P_{Z_r} = P_Z + P_{M_Z Y_o}$, we have $A^* = A - P_B X = A - B(B'B)^{-1}B'X$, so $D = (B'B)^{-1}B'X$. Thus $P_{(A,B)} = P_{(A^*,B)} = P_{(P_Z X, M_Z Y_o)}$ giving

$$\hat{u}_r' P_{Z_r} \hat{u}_r - \hat{u}' P_Z \hat{u} = y'(P_{(P_{Z_r} X, M_Z Y_o)} - P_{P_{Z_r} X})y. \tag{2.42}$$

Hence, in addition to H_o, S_o establishes yet another hybrid form combining elements of both W_o and D_o. However, when $L = K$ then $S_o = D_o$, due to $P_Z \hat{u} = 0$.

2.7 Concerns for practitioners

The foregoing subsections demonstrate that all five available archetypal statistics W_o, D_o, T_o, H_o and S_o for testing the orthogonality of a subset of the potentially endogenous regressors basically just differ regarding the way in which the expresssions they are based on are scaled with respect to estimates of σ^2. The two tests H_o and S_o show a hybrid nature in this respect, because their most natural implementation requires two different σ^2 estimates, which may lead to negative test outcomes. In

[11] Ruud (2000, p.582) proves this just for the special case $K_e = 0$. Newey (1985, p.238), Baum et al. (2003, p.23 and formula 55) and Hayashi (2002) mention equivalence for $K_e \geq 0$, but do not provide a proof.

addition to that, H_o has the drawback that it involves a generalized inverse, whereas calculation of all the others is rather straight-forward.[12]

Similar differences and correspondences will carry over to more general models, which would require either robust variance estimation or GMM, see Newey (1985) and Ahn (1997). Although of no concern asymptotically, these differences may have major consequences in finite samples, thus practitioners are in need of clues which implementations should be preferred.[13] Therefore, in the remainder of this study, we will examine the performance in finite samples of all these six archetypal tests. First, we will examine whether they procure acceptable size control, possibly after a simple degrees of freedom correction or after bootstrapping. Next, only for those variants that pass this requirement we will perform some power calculations.

3 Earlier Monte Carlo designs and results

In the literature the actual rejection frequencies of tests on the independence between regressors and disturbances have been examined by simulation only for situations where all possibly endogenous regressors are tested jointly, hence $K_e = 0$. To our knowledge, sub-set orthogonalty tests have not been examined yet.

Wu (1974) was the first to design a simulation study in which he examined the four tests suggested in Wu (1973). He made substantial efforts, both analytically and experimentally, to assess the parameters and model characteristics which actually determine the distribution of the test statistics and their power curves. His focus is on the case where there is one possibly endogenous regressor ($K_o = 1$), an intercept and one other included exogenous regressor ($L_1 = 2$) and two external instruments ($L_2 = 2$), giving a degree of overidentification of 1. All disturbances are assumed normal, all exogenous regressors are mutually orthogonal and all consist of elements equal to either 1, 0, or -1, whereas all instruments have coefficient 1 in the reduced form. Wu demonstrates that all considered test statistics are functions of statistics that follow Wishart distributions which are invariant with respect to the values of the structural coefficients of the equation of interest. The effects of changing the degree of simultaneity and of changing the joint strength of the external instruments are examined. Because the design is rather inflexible regarding varying the explanatory

[12] It is not obvious why Pesaran and Smith (1990, p.49,55) mention that they find T_o a computationally more attractive statistic than W_o. Apart from H_o all discrepancy based test statistics are very easy to compute. However, T_o is the only one that strictly applies a standard procedure (Wald) to testing zero restrictions in an auxiliary regression, which eases its use by standard software packages. On the other hand Baum et al. (2003, p.26) characterize tests like T_o as "computationally expensive and practically cumbersome", which we find far fetched too.

[13] Under the heading of "Regressor Endogeneity test" EViews 8.1 presents statistic S_o where for both σ^2 estimates an $n - K$ degrees of freedom correction is used, like it does for the J statistic. In Stata 13 the "hausman" command calculates H_o by default and offers the possibility to calculate W_o and D_o. The degrees of freedom reported is the rank of the estimated variance of the discrepancy vector. In case of H_o this is not correct. It is possible to overwrite the degrees of freedom by an additional command. The popular package "ivreg2" only reports D_o with the correct degrees of freedom.

part of the reduced form, no separate attention is paid to the effects of multicollinearity of the regressors on the rejection probabilities, nor to the effects of weakness of individual instruments. Although none of the tests examined is found to be superior under all circumstances, test T_2, which is exact under normality and generalized as T_o in (2.28), is found to be the preferred one. Its power increases with the absolute value of the degree of simultaneity, with the joint strength of the instruments and with the sample size.

Nakamura and Nakamura (1985) examine a design where $K_e = 0$, $K_o = 1$, $L_1 = 2$, $L_2 = 3$ and all instruments are mutually independent standard normal. The structural equation disturbances u and the reduced form disturbances v are IID normal with variances σ_u^2 and σ_v^2 respectively and correlation ρ. They focus on the case where all coefficients in the structural equation and in the reduced form equation for the possibly endogenous regressor are unity. Given the fixed parameters the distribution of the test statistic T_2 now depends only on the values of ρ^2, σ_u^2 and σ_v^2. Attention is drawn to the fact that the power of an endogeneity test and its interpretation differs depending on whether the test is used to signal: (a) the degree of simultaneity expressed as ρ, (b) the simultaneity expressed as the covariance $\delta = \rho\sigma_u\sigma_v$, or (c) the extent of the asymptotic bias of OLS (which in their design is also determined just by ρ, σ_u^2 and σ_v^2). When testing (a) a natural choice of the nuisance parameters (which are kept fixed when ρ is varied to obtain a power curve) are σ_u and σ_v. However, when testing (b) or (c) ρ, σ_u and σ_v cannot all be chosen independently. The study shows that, although the power of test T_2 does increase for increasing values of ρ^2 while keeping σ_u and σ_v constant, it may decrease for increasing asymptotic OLS bias. Therefore, test T_2 is not very suitable for signaling the magnitude of OLS bias. In this design $\sigma_v^2 = 5(1 - R^2)/R^2$, where R^2 is the population coefficient of determination of the reduced form equation for the possibly endogenous regressor. The joint strength of the instruments is a simple function of R^2 and hence of σ_v. Again, due to the fixed values of the reduced form coefficients the effects of weakness of individual instruments or of multicollinearity cannot be examined from this design.

The study by Kiviet (1985) demonstrates that in models with a lagged dependent explanatory variable the actual type I error probability of test T_2 may deviate substantially from the chosen nominal level. Then high rejection frequencies under the alternative have little or no meaning.[14] In the present study we will stick to static cross-section type models.

Thurman (1986) performs a small scale Monte Carlo simulation of just 100 replications on a specific two equation simultaneous model using empirical data for the exogenous variables from which he concludes that Wu-Hausman tests may have substantial power under particular parametrizations and none under others.

Chmelarova and Hill (2010) focus on pre-test estimation based on test T_2 (for $K_o = 1$, $L_1 = 2$, $L_2 = 1$) and two other forms of contrast based tests which use an improper number of degrees of freedom[15]. Their Monte Carlo design is very restricted,

[14] Because we could not replicate some of the presented figures for the case of strong instruments, we plan to re-address the analysis of DWH type tests in dynamic models in future work.

[15] This may occur when standard software is employed based on a naive implementation of the Hausman test. Practitioners should be adviced never to use these standard options. Confusion about

because the possibly endogenous regressor and the exogenous regressor (next to the constant) are uncorrelated, so multicollinearity does not occur, which makes the DGP unrealistic. Moreover, all coefficients in the equation of interest are kept fixed and are such that the signal to noise ratio is always 1. Therefore, the inconsistency of OLS is relatively large (and in fact equal to the simultaneity correlation coefficient ρ). Because the sample size is not varied and neither is the instrument strength parameter[16] the results do not allow to form an opinion on how effective the T_2 test is to diagnose simultaneity.

Jeong and Yoon (2010) present a study in which they examine by simulation what the rejection probability of the Hausman test is when an instrument is employed which is actually correlated with the disturbances. Also for the sub-set tests to be examined here the situation seems of great practical relevance that they might be implemented while using some variable(s) as instruments which are in fact endogenous. In our Monte Carlo experiments we will cover such situations, but we do not find the design as used by Jeong and Yoon, in which the endogeneity/exogeneity status of variables depends on the sample size very useful.

4 A more comprehensive Monte Carlo design

To examine the differences between the various sub-set tests regarding their type I and II error probabilities in finite samples we want to lay out a Monte Carlo design which is less restrictive than those just reviewed. It should allow to represent the major characteristics of data series and their relationships as faced in empirical work, while avoiding the imposition of awkward restrictions on the nuisance parameter space. Instead of picking particular values for the coefficients and further parameters in a simple DGP, and check whether or not this leads to covering empirically relevant cases, we choose to approach this design problem from the opposite direction.

4.1 The simulated data generating process

Model (2.1) is specialized in our simulations to cases where $K = 3$ with

$$y = \beta_1 \iota + \beta_2 y^{(2)} + \beta_3 y^{(3)} + u, \tag{4.1}$$

$$y^{(2)} = \pi_{21} \iota + \pi_{22} z^{(2)} + \pi_{23} z^{(3)} + v^{(2)}, \tag{4.2}$$

$$y^{(3)} = \pi_{31} \iota + \pi_{32} z^{(2)} + \pi_{33} z^{(3)} + v^{(3)}, \tag{4.3}$$

where ι is an $n \times 1$ vector consisting of ones. Depending on the disturbance correlations the variables $y^{(2)}$ and $y^{(3)}$ can be either endogenous or exogenous. We will examine situations where L_1 is either 1 or 2. When $L_1 = 1$, $Z_1 = \iota$ and $Z_2 = (z^{(2)}, z^{(3)})$ so

the correct number of restrictions is avoided by using a test based on estimator contrast, which can be obtained by running either the relevant auxiliary regression or the two alternative regressions and obtain the Sargan incremental test.

[16] If the effects of a weaker instrument had been checked the simulation estimates of the moments of IV (which do not exist, because the model is just identified) would have gone astray.

$L_2 = 2$ and $L = 3$, whereas $Y = (y^{(2)}, y^{(3)})$, so $K_o + K_e = 2 = K - L_1$. Now K_o is either 1 (for sub-set tests, with $K_e = 1$) or 2 (for full-set tests, since $K_e = 0$). In case $K_o = 1$ then Y_o is either $y^{(3)}$ or $y^{(2)}$, and when $K_o = 2$ then $Y_o = (y^{(2)}, y^{(3)})$. In all these cases $L = K$, so the single simultaneous equation (4.1) is just identified according to the order condition under the unrestrained alternative hypothesis. When $L_1 = 2$ then Z_1 is either $(\iota, y^{(2)})$ or $(\iota, y^{(3)})$, with Y_o either $y^{(3)}$ or $y^{(2)}$, so $K_o = 1$ and $K_e = 0$. Thus here we just have full-set tests, and because $Z_2 = (z^{(2)}, z^{(3)})$ again, $L_2 = 2$ giving $L = 4$. Hence, when $L_1 = 2$ there is mild overidentification because $L - K = 1$.

As the statistics to be analyzed are all invariant regarding the values of the intercepts, these are all set equal to zero, thus $\beta_1 = \pi_{21} = \pi_{31} = 0$. For the vectors $z^{(2)}$ and $z^{(3)}$ two arbitrary mutually independent $IID(0,1)$ series were drawn and kept fixed over all replications. In fact, we rescaled them such that their sample mean and variance, and also their sample covariance correspond to the population values 0, 1 and 0 respectively. To allow for simultaneity of both $y^{(2)}$ and $y^{(3)}$, as well as for any value of the correlation between the reduced form disturbances $v^{(2)}$ and $v^{(3)}$, these disturbances were generated as

$$v^{(2)} = \eta^{(2)} + \gamma_2 u \text{ and } v^{(3)} = \eta^{(3)} + \kappa\eta^{(2)} + \gamma_3 u, \tag{4.4}$$

where the series u_i, $\eta_i^{(2)}$ and $\eta_i^{(3)}$ are mutually independent zero mean IID series (for $i = 1, ..., n$). Without loss of generality, we chose $\sigma_u^2 = 1$. Scaling the variances of the potentially endogenous regressors simplifies the model even further, again without loss of generality. This scaling is innocuous, because it can be compensated by the chosen values for β_2 and β_3. We realized $\sigma^2_{y^{(2)}} = \sigma^2_{y^{(3)}} = 1$ by choosing appropriate values for $\sigma^2_{\eta^{(2)}} > 0$ and $\sigma^2_{\eta^{(3)}} > 0$ as follows. For the variance of the IID series for the reduced form disturbances and for the possibly endogenous explanatory variables we find

$$\begin{aligned} \sigma^2_{v^{(2)}} &= \sigma^2_{\eta^{(2)}} + \gamma_2^2, & \sigma^2_{y^{(2)}} &= \pi_{22}^2 + \pi_{23}^2 + \sigma^2_{v^{(2)}} = 1, \\ \sigma^2_{v^{(3)}} &= \sigma^2_{\eta^{(3)}} + \kappa^2\sigma^2_{\eta^{(2)}} + \gamma_3^2, & \sigma^2_{y^{(3)}} &= \pi_{32}^2 + \pi_{33}^2 + \sigma^2_{v^{(3)}} = 1. \end{aligned} \tag{4.5}$$

This requires

$$\sigma^2_{\eta^{(2)}} = 1 - \pi_{22}^2 - \pi_{23}^2 - \gamma_2^2 > 0 \text{ and } \sigma^2_{\eta^{(3)}} = 1 - \pi_{32}^2 - \pi_{33}^2 - \kappa^2\sigma^2_{\eta^{(2)}} - \gamma_3^2 > 0. \tag{4.6}$$

Also, when both $y^{(2)}$ and $y^{(3)}$ are endogenous, fulfillment of the rank condition for identification requires

$$\pi_{22}\pi_{33} \neq \pi_{23}\pi_{32}. \tag{4.7}$$

However, as we will see, more restrictions than those given by (4.7) and (4.6) should be respected when we consider further consequences of a choice of particular values for the nine remaining DGP parameters

$$\{\gamma_2, \gamma_3, \kappa, \pi_{22}, \pi_{23}, \pi_{32}, \pi_{33}, \beta_2, \beta_3\}. \tag{4.8}$$

4.2 Simulation design parameter space

Assigning a range of reasonable values to the nine DGP parameters is cumbersome as it is not immediately obvious what model characteristics they imply. Therefore, we now first define econometrically meaningful design parameters. These are functions of the DGP parameters, and we will invert these functions in order to find solutions for the parameters of the DGP in terms of the chosen design parameter values. Since the DGP is characterized by nine parameters, we should define nine variation free design parameters as well. However, their relationships will be such, that this will not automatically imply the existence nor the uniqueness of solutions.

Two obvious design parameters are the degree of simultaneity in $y^{(2)}$ and $y^{(3)}$, given by

$$\rho_j = Cov(y_i^{(j)}, u_i)/(\sigma_{y^{(j)}}\sigma_u) = \gamma_j, \ j = 2, 3. \tag{4.9}$$

Hence, by choosing $\sigma^2_{y^{(2)}} = \sigma^2_{y^{(3)}} = 1$, the degree of simultaneity in $y^{(j)}$ is directly controlled by γ_j for $j = 2, 3$, and it implies two more inequality restrictions, namely

$$|\gamma_j| < 1, \ j = 2, 3. \tag{4.10}$$

Another design parameter is a measure of multicollinearity between $y^{(2)}$ and $y^{(3)}$ given by the correlation

$$\rho_{23} = \pi_{22}\pi_{32} + \pi_{23}\pi_{33} + \kappa(1 - \pi_{22}^2 - \pi_{23}^2 - \gamma_2^2) + \gamma_2\gamma_3, \tag{4.11}$$

implying yet another restriction

$$\left|\pi_{22}\pi_{32} + \pi_{23}\pi_{33} + \kappa(1 - \pi_{22}^2 - \pi_{23}^2 - \gamma_2^2) + \gamma_2\gamma_3\right| < 1. \tag{4.12}$$

Further characterizations relevant from an econometric perspective are the marginal strength of instrument $z^{(2)}$ for $y^{(j)}$ and the joint strength of $z^{(2)}$ and $z^{(3)}$ for $y^{(j)}$, which are established by the (partial) population coefficients of determination

$$R^2_{j;z2} = \pi^2_{j2} \text{ and } R^2_{j;z23} = \pi^2_{j2} + \pi^2_{j3}, \ j = 2, 3. \tag{4.13}$$

In the same vain, and completing the set of nine design parameters, are two similar characterizations of the fit of the equation of interest. Because the usual R^2 gives complications under simultaneity, we focus on its reduced form equation

$$\begin{aligned} y = {} & (\beta_2\pi_{22} + \beta_3\pi_{32})\, z^{(2)} + (\beta_2\pi_{23} + \beta_3\pi_{33})\, z^{(3)} \\ & + (\beta_2 + \beta_3\kappa)\, \eta^{(2)} + \beta_3\eta^{(3)} + (1 + \beta_2\gamma_2 + \beta_3\gamma_3)\, u. \end{aligned} \tag{4.14}$$

This yields

$$\begin{aligned} \sigma_y^2 = {} & (\beta_2\pi_{22} + \beta_3\pi_{32})^2 + (\beta_2\pi_{23} + \beta_3\pi_{33})^2 \\ & + (\beta_2 + \beta_3\kappa)^2\, \sigma^2_{\eta^{(2)}} + \beta_3^2\sigma^2_{\eta^{(3)}} + (1 + \beta_2\gamma_2 + \beta_3\gamma_3)^2, \end{aligned} \tag{4.15}$$

and in line with (4.13) we then have

$$\begin{aligned} & R^2_{1;z2} = (\beta_2\pi_{22} + \beta_3\pi_{32})^2 / \sigma_y^2 \text{ and} \\ & R^2_{1;z23} = [(\beta_2\pi_{22} + \beta_3\pi_{32})^2 + (\beta_2\pi_{23} + \beta_3\pi_{33})^2]/\sigma_y^2. \end{aligned} \tag{4.16}$$

The 9-dimensional design parameter space is given now by

$$\{\rho_2, \rho_3, \rho_{23}, R^2_{2;z2}, R^2_{2;z23}, R^2_{3;z2}, R^2_{3;z23}, R^2_{1;z2}, R^2_{1;z23}\}. \tag{4.17}$$

The first three of these parameters have domain $(-1, +1)$ and the six R^2 values have to obey the restrictions

$$0 \leq R^2_{j;z2} \leq R^2_{j;z23} < 1,\ j = 1, 2, 3. \tag{4.18}$$

However, without loss of generality we can further restrict the domain of the nine design parameters, due to symmetry of the DGP with respect to: (a) the two regressors $y^{(2)}$ and $y^{(3)}$ in (4.1), (b) the two instrumental variables $z^{(2)}$ and $z^{(3)}$, and (c) implications which follow when all random variables are drawn from distributions with a symmetric density function.

Due to (a) we may just consider cases where

$$\rho_2^2 \geq \rho_3^2. \tag{4.19}$$

So, if one of the two regressors will have a more severe simultaneity coefficient, it will always be $y^{(2)}$. Due to (b) we will limit ourselves to cases where $\pi_{22}^2 \geq \pi_{23}^2$. Hence, if one of the instruments for $y^{(2)}$ is stronger than the other, it will always be $z^{(2)}$. On top of (4.18) this implies

$$R^2_{2;z2} \geq 0.5 R^2_{2;z23}. \tag{4.20}$$

If (c) applies, we may restrict ourselves to cases where next to particular values for (γ_2, γ_3), we do not also have to examine $(-\gamma_2, -\gamma_3)$. This is achieved by imposing $\rho_2 + \rho_3 \geq 0$. In combination with (4.19) this leads to

$$1 > \rho_2 \geq |\rho_3| \geq 0. \tag{4.21}$$

Solving the DGP parameters in terms of the design parameters can now be achieved as follows. In a first stage we can easily solve 7 of the 9 parameters, namely

$$\left.\begin{aligned} &\gamma_j = \rho_j \\ &\pi_{j2} = d_{j2} \left|(R^2_{j;z2})^{1/2}\right|,\ d_{j2} = -1, +1 \\ &\pi_{j3} = d_{j3} \left|(R^2_{j;z23} - R^2_{j;z2})^{1/2}\right|,\ d_{j3} = -1, +1 \end{aligned}\right\} j = 2, 3. \tag{4.22}$$

With (4.11) these give

$$\kappa = (\rho_{23} - \pi_{22}\pi_{32} - \pi_{23}\pi_{33} - \gamma_2\gamma_3)/(1 - \pi_{22}^2 - \pi_{23}^2 - \gamma_2^2). \tag{4.23}$$

Thus, for a particular case of chosen design parameter values, obeying the inequalities (4.18) through (4.21), we may obtain 2^4 solutions from (4.22) and (4.23) for the DGP parameters. However, some of these may be inadmissible, if they do not fulfill the requirements (4.7) and (4.6). Moreover, we will show that not all of these 2^4 solutions necessarily lead to unique results on the distribution of the test statistics.

Finally, the remaining two parameters β_2 and β_3 can be solved from the pair of nonlinear equations

$$\left.\begin{aligned}(1-R^2_{1;z2})\left(\beta_2\pi_{22}+\beta_3\pi_{32}\right)^2 &= R^2_{1;z2}[(\beta_2\pi_{23}+\beta_3\pi_{33})^2 \\ &\quad +(1+\beta_2\gamma_2+\beta_3\gamma_3)^2+\beta_3^2\sigma^2_{\eta^{(3)}}+(\beta_2+\beta_3\kappa)^2\,\sigma^2_{\eta^{(2)}}], \\ (1-R^2_{1;z23})[(\beta_2\pi_{22}+\beta_3\pi_{32})^2+(\beta_2\pi_{23}+\beta_3\pi_{33})^2] &= R^2_{1;z23}[(1+\beta_2\gamma_2+\beta_3\gamma_3)^2 \\ &\quad +\beta_3^2\sigma^2_{\eta^{(3)}}+(\beta_2+\beta_3\kappa)^2\,\sigma^2_{\eta^{(2)}}].\end{aligned}\right\} \tag{4.24}$$

Both these equations represent particular conic sections, specializing into either ellipses, parabolas or hyperbolas, implying that there may be zero up to eight solutions. However, it is easy to see that five of the archetypal sub-set test statistics are invariant with respect to β. Note that $\hat{u}=[I-X(X'P_ZX)^{-1}X'P_Z](X\beta+u)=[I-X(X'P_ZX)^{-1}X'P_Z]u$ and $\hat{u}_r=[I-X(X'P_{Z_r}X)^{-1}X'P_{Z_r}]u$ are invariant with respect to β, thus so are $\hat{\sigma}^2$ and $\hat{\sigma}^2_r$ and the expressions in (2.35), and therefore the statistic S_o and H_o are. And $\ddot{\sigma}^2$ is invariant with respect to β too, because $y-X\hat{\beta}-M_ZY_o\hat{\xi}=\hat{u}-M_ZY_o\hat{\xi}$ is, as follows from (2.22) and (2.23). Moreover, because $\hat{\xi}$ is invariant with respect to β their numerator (like their respective denominators) and thus the three test statistics W_o, D_o and T_o are.[17] Therefore, $R^2_{1;z2}$ and $R^2_{1;z23}$ do not really establish nuisance parameters (except perhaps for WD_o), reducing the dimensionality of the nuisance parameter space to 7. Without loss of much generality we may therefore set $\beta_2=\beta_3=0$ in the simulated DGP's.

When (c) applies, not all 16 permutations of the signs of the four reduced form coefficients lead to unique results for the test statistics, because of the following. If the sign of all elements of $y^{(2)}$ and (or) $y^{(3)}$ is changed, this means that in the general formulas the matrix X is replaced by XJ, where J is a $K\times K$ diagonal matrix with diagonal elements $+1$ or -1, for which $J=J'=J^{-1}$. It is easily verified that such a transformation has no effect on the quadratic forms in y which constitute the six test statistics, nor that of the projection matrices used in the three different estimators of σ^2. So, when changing the sign of all reduced form coefficients, and at the same time the sign of all the elements of the vectors u, $\eta^{(2)}$ and $\eta^{(3)}$, the same test statistics are found, whereas the simultaneity and multicollinearity are still the same. This reduces the 16 possible permutations to 8, which we achieve by choosing $d_{22}=1$. From the remaining 8 permutations four different couples yield similar ρ_{23} and κ values. We keep the four permutations which genuinely differ by choosing $d_{23}=1$, and will give explicit attention to the four distinct cases

$$(d_{22},d_{23},d_{32},d_{33})=\begin{cases}(1,1,1,1)\\(1,1,-1,1)\\(1,1,1,-1)\\(1,1,-1,-1),\end{cases} \tag{4.25}$$

when we generate the disturbances from a symmetric distribution, which at this stage we will.

[17] Wu (1974) finds this invariance result too, but his proof suggests that it is a consequence of normality of all the disturbances, whereas it holds more generally.

For the design parameters we shall choose various interesting combinations from

$$\left.\begin{array}{rl} \rho_2 \in & \{0, .2, .5\} \\ \rho_3 \in & \{-.5, -.2, 0, .2, .5\} \\ \rho_{23} \in & \{-.5, -.2, 0, .2, .5\} \\ R^2_{j;z2} \in & \{.01, .1, .2, .3\} \\ R^2_{j;z23} \in & \{.02, .1, .2, .4, .5, .6\} \end{array}\right\} \tag{4.26}$$

in as far as they satisfy the restrictions (4.18) through (4.21), provided they obey also the admissibility restrictions given by (4.7), (4.6) and (4.12).

5 Simulation findings on rejection probabilities

In each of the R replications in the simulation study, new independent realizations are drawn on u, $\eta^{(2)}$ and $\eta^{(3)}$. All the five archetypal test statistics W_o, D_o, T_o, H_o and S_o will in principle be examined for five different implementations. Two of these allow genuine sub-set tests. They test the endogeneity of either $y^{(2)}$ (then denoted as W^2, D^2, etc.) or of $y^{(3)}$ (denoted W^3, etc.) assuming the other regressor to be endogenous. Three implementations constitute full set tests (which does not mean the full set of regressors, but refers to the full set of regressors deemed to be endogenous). They are denoted W_3^2, etc. (when the endogeneity of $y^{(2)}$ is tested and $y^{(3)}$ is treated as exogenous) or W_2^3, etc. (when $y^{(2)}$ is treated as exogenous) or W^{23}, etc. (when the endogeneity of both $y^{(2)}$ and $y^{(3)}$ is tested). The behavior under both the null and the alternative hypothesis will be investigated. The full-set tests are included to better appreciate the special nature of the more subtle sub-set tests. Since $D^2 = S^2$, $D^3 = S^3$ and $D^{23} = S^{23}$ because $L = K$ the tables will just mention the D variant.

Every replication it is checked whether or not the null hypothesis is rejected by test statistic Υ, where Υ is any of the tests indicated above. From this we obtain the Monte Carlo estimate

$$\overrightarrow{p}_\Upsilon = \frac{1}{R}\textstyle\sum_{r=1}^{R}\mathbb{I}\left(\Upsilon^{(r)} > \Upsilon^c\left(\alpha\right)\right). \tag{5.1}$$

Here $\mathbb{I}\left(.\right)$ is the indicator function that takes value one when its argument is true and zero when it is not. We take the standard form of the test statistics in which $\Upsilon^c(\alpha)$ is the α-level critical value of the χ^2 distribution (with either 1 or 2 degrees of freedom) and in which σ^2 estimates have no degrees of freedom correction.

The Monte Carlo estimator $\overrightarrow{p}_\Upsilon$ estimates the actual rejection probability of asymptotic test procedure Υ. When H^0 is true it estimates the actual type I error probability (at nominal level α) and when H^0 is false $1-\overrightarrow{p}_\Upsilon$ estimates the type II error probability, whereas $\overrightarrow{p}_\Upsilon$ is then a (naive, when there are size distortions) estimator of the power function of the test in one particular argument (defined by the specific case of values of the design and matching DGP parameters). Estimator $\overrightarrow{p}_\Upsilon$ follows the binomial distribution and has standard errors given by $SE(\overrightarrow{p}_\Upsilon) = [\overrightarrow{p}_\Upsilon(1-\overrightarrow{p}_\Upsilon)/R]^{1/2}$. For R large, a 99.75% confidence interval for the true rejection probability is

$$\mathrm{CI}_{99.75\%} = [\overrightarrow{p}_\Upsilon - 3 * SE(\overrightarrow{p}_\Upsilon),\ \overrightarrow{p}_\Upsilon + 3 * SE(\overrightarrow{p}_\Upsilon)]. \tag{5.2}$$

We choose $R = 10000$, examine all endogeneity tests at the nominal significance level of 5%, and report results for sample size equal to $n = 40$ (but also checked the major results at $n = 100$). Note that the boundary values for determining whether the actual type I error probability of these asymptotic tests differs at this particular small sample size significantly (at the very small level of .25%) from the nominal value 5% are .043 and .057 respectively. Versions of tests that are found to have poor size control in the initial tables will be skipped from further examination.

5.1 At least one exogenous regressor

In this subsection we examine cases where either both regressors $y^{(2)}$ and $y^{(3)}$ are actually exogenous or just $y^{(3)}$ is exogenous. Hence, for particular implementations of the sub-set and full-set tests on endogeneity the null hypothesis is true, but for some it is false. In fact, it is always true for the sub-set tests on $y^{(3)}$ in the cases of this subsection. We present a series of tables containing estimated rejection probabilities and each separate table focusses on a particular setting regarding the strength of the instruments. Every case consists of potentially four subcases; "a" stands for $(d_{32}, d_{33}) = (1, 1)$, "b" for $(d_{32}, d_{33}) = (-1, 1)$, "c" for $(d_{32}, d_{33}) = (1, -1)$ and "d" for $(d_{32}, d_{33}) = (-1, -1)$. When both instruments have similar strength for $y^{(2)}$ and also (but probably stronger or weaker) for $y^{(3)}$ the identification condition requires $d_{32} \neq d_{33}$. Then only two of the four combinations (4.25) are feasible so that every case just consists of the two subcases "b" and "c".

In Table 1 we consider cases with mildly strong instruments. In the first five cases both $y^{(2)}$ and $y^{(3)}$ are exogenous whereas the degree of multicollinearity changes. So in the first ten rows of the table, for all five distinct implementations of the five different test statistics examined, the null hypothesis is true. Because $y^{(2)}$ and $y^{(3)}$ are parametrized similarly here, the two sub-set test implementations are actually equivalent. The minor differences in rejection probabilities stem from random variation, both in the disturbances and in the single realizations of the instruments. The same holds for the two full-set implementations with one degree of freedom. For all implementations over the first five cases (both "b" and "c") D_o and S_o show acceptable size control, whereas W_o tends to underreject, whilst T_o overrejects and H_o does both. The sub-set version of W_o gets worse under multicollinearity (irrespective of the sign of ρ_{23}), whereas multicollinearity increases the type I error probability of the full-set W_o tests. Both D_o and T_o seem unaffected by multicollinearity for these cases.

When $y^{(2)}$ is made mildly endogenous, as in cases 6-10, the null hypothesis is still true for the sub-set tests W^3, etc. Their type I error probability seems virtually unaffected by the actual values of ρ_2 and ρ_{23}. For the sub-set tests W^2, etc. the null hypothesis is false. Due to their differences in type I error probability we cannot conclude much about power yet, but that they have some and that it is virtually unaffected by ρ_{23} is clear. The next three columns demonstrate that it is essential that a full-set test comprises all endogenous regressors, because if it does not the test may falsely diagnose endogeneity of an exogenous regressor (but by a reasonably low probability when the regressors are hardly correlated). The next implementation reported, in which the exogeneity of $y^{(3)}$ is exploited, demonstrate that in this case

Table 1: One endogenous regressor and mildly strong instruments:

$R^2_{2;z2} = 0.20,\ R^2_{2;z23} = 0.40,\ R^2_{3;z2} = 0.20,\ R^2_{3;z23} = 0.40$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	H^3	W^2	D^2	T^2	H^2	W_2^3	D_2^3	T_2^3	H_2^3	S_2^3	W_3^2	D_3^2	T_3^2	H_3^2	S_3^2	W^{23}	D^{23}	T^{23}	H_c^{23}
1b	0.0	0.0	0.0	0.032	0.054	0.065	0.022	0.033	0.055	0.064	0.023	0.037	0.046	0.070	0.033	0.049	0.038	0.047	0.069	0.034	0.049	0.021	0.042	0.084	0.011
1c	0.0	0.0	0.0	0.032	0.053	0.062	0.023	0.032	0.051	0.062	0.023	0.040	0.047	0.068	0.035	0.048	0.038	0.048	0.069	0.034	0.049	0.021	0.041	0.085	0.011
2b	0.0	0.0	-0.2	0.029	0.054	0.064	0.019	0.030	0.053	0.063	0.022	0.043	0.048	0.074	0.039	0.050	0.043	0.048	0.070	0.040	0.049	0.022	0.042	0.083	0.013
2c	0.0	0.0	-0.2	0.030	0.052	0.061	0.020	0.030	0.052	0.061	0.021	0.042	0.048	0.070	0.038	0.049	0.043	0.047	0.069	0.040	0.049	0.022	0.042	0.086	0.012
3b	0.0	0.0	0.2	0.031	0.053	0.064	0.020	0.032	0.054	0.063	0.022	0.041	0.047	0.069	0.037	0.048	0.042	0.048	0.070	0.039	0.049	0.022	0.042	0.085	0.012
3c	0.0	0.0	0.2	0.030	0.052	0.063	0.022	0.029	0.050	0.062	0.021	0.040	0.045	0.067	0.037	0.046	0.042	0.048	0.069	0.040	0.049	0.022	0.043	0.084	0.012
4b	0.0	0.0	-0.5	0.021	0.051	0.060	0.013	0.024	0.051	0.062	0.016	0.055	0.048	0.072	0.093	0.048	0.053	0.048	0.070	0.092	0.047	0.026	0.043	0.082	0.030
4c	0.0	0.0	-0.5	0.025	0.056	0.067	0.015	0.025	0.057	0.063	0.015	0.054	0.048	0.070	0.091	0.047	0.053	0.046	0.071	0.092	0.046	0.025	0.042	0.084	0.029
5b	0.0	0.0	0.5	0.025	0.057	0.069	0.017	0.024	0.055	0.067	0.016	0.054	0.048	0.071	0.094	0.047	0.053	0.048	0.070	0.090	0.047	0.026	0.045	0.087	0.027
5c	0.0	0.0	0.5	0.023	0.053	0.062	0.017	0.022	0.052	0.062	0.015	0.052	0.047	0.068	0.092	0.046	0.053	0.047	0.071	0.093	0.046	0.025	0.042	0.084	0.028
6b	0.2	0.0	0.0	0.031	0.056	0.065	0.020	0.124	0.179	0.201	0.096	0.038	0.048	0.071	0.032	0.051	0.136	0.159	0.206	0.124	0.163	0.063	0.118	0.195	0.037
6c	0.2	0.0	0.0	0.031	0.054	0.061	0.021	0.125	0.175	0.197	0.097	0.037	0.047	0.069	0.033	0.050	0.139	0.162	0.207	0.126	0.165	0.064	0.119	0.193	0.038
7b	0.2	0.0	-0.2	0.028	0.055	0.062	0.019	0.112	0.165	0.196	0.090	0.073	0.083	0.113	0.069	0.085	0.172	0.189	0.243	0.164	0.193	0.085	0.141	0.225	0.053
7c	0.2	0.0	-0.2	0.030	0.053	0.061	0.020	0.119	0.173	0.197	0.093	0.070	0.078	0.109	0.064	0.080	0.173	0.190	0.242	0.162	0.194	0.083	0.141	0.224	0.051
8b	0.2	0.0	0.2	0.031	0.054	0.063	0.021	0.118	0.172	0.200	0.093	0.072	0.080	0.113	0.066	0.083	0.174	0.193	0.244	0.164	0.197	0.085	0.145	0.228	0.052
8c	0.2	0.0	0.2	0.030	0.053	0.062	0.020	0.118	0.169	0.196	0.095	0.072	0.081	0.110	0.067	0.083	0.174	0.192	0.242	0.166	0.194	0.089	0.142	0.225	0.056
9b	0.2	0.0	-0.5	0.022	0.052	0.062	0.016	0.092	0.144	0.184	0.076	0.700	0.682	0.741	0.778	0.672	0.757	0.739	0.791	0.831	0.734	0.600	0.637	0.742	0.558
9c	0.2	0.0	-0.5	0.026	0.057	0.066	0.019	0.096	0.151	0.190	0.077	0.702	0.685	0.741	0.783	0.678	0.759	0.745	0.795	0.833	0.740	0.609	0.643	0.750	0.561
10b	0.2	0.0	0.5	0.026	0.058	0.066	0.019	0.096	0.150	0.189	0.078	0.704	0.688	0.743	0.782	0.679	0.762	0.744	0.798	0.832	0.738	0.613	0.650	0.749	0.568
10c	0.2	0.0	0.5	0.024	0.053	0.062	0.017	0.095	0.146	0.185	0.077	0.705	0.686	0.745	0.785	0.678	0.765	0.747	0.801	0.833	0.743	0.614	0.647	0.747	0.568
11b	0.5	0.0	0.0	0.027	0.062	0.061	0.015	0.816	0.867	0.886	0.768	0.041	0.051	0.075	0.036	0.069	0.822	0.844	0.882	0.806	0.849	0.642	0.788	0.863	0.500
11c	0.5	0.0	0.0	0.028	0.059	0.058	0.014	0.816	0.868	0.887	0.770	0.041	0.050	0.072	0.036	0.066	0.829	0.850	0.887	0.813	0.854	0.642	0.785	0.861	0.499
12b	0.5	0.0	-0.2	0.026	0.060	0.057	0.017	0.768	0.814	0.853	0.737	0.322	0.345	0.411	0.304	0.373	0.926	0.931	0.950	0.921	0.932	0.812	0.895	0.939	0.706
12c	0.5	0.0	-0.2	0.028	0.060	0.056	0.017	0.766	0.814	0.857	0.735	0.320	0.341	0.406	0.307	0.367	0.927	0.931	0.950	0.923	0.932	0.819	0.893	0.940	0.704
13b	0.5	0.0	0.2	0.029	0.063	0.060	0.020	0.767	0.818	0.860	0.736	0.327	0.349	0.412	0.312	0.373	0.928	0.932	0.950	0.925	0.933	0.821	0.896	0.939	0.708
13c	0.5	0.0	0.2	0.026	0.058	0.059	0.017	0.773	0.816	0.858	0.742	0.329	0.350	0.418	0.313	0.379	0.928	0.932	0.951	0.924	0.933	0.817	0.895	0.941	0.706

full-set tests do a much better job in detecting the endogenous nature of $y^{(2)}$ than the sub-set tests, provided there is (serious) multicollinearity. Here the full-set tests have the advantage of using an extra valid instrument. The effects of multicollinearity can be explained as follows. Using the notation of the more general setup and auxiliary regression (2.19), the sub-set (full-set) tests test here the significance of the regressors $P_Z Y_o$ ($P_{Z^*} Y_o$) in a regression already containing $P_{Z_r} X$ ($P_{Z_r^*} X = X$), where $Z^* = (Z, Y_e)$ and $Z_r^* = (Z_r, Y_e)$. Regarding the sub-set test it is obvious that, because the space spanned by $P_{Z_r} X = (P_{Z_r} Y_e, Y_o, Z_1)$ does not change when Y_e and Y_o are more or less correlated, the significance test of $P_Z Y_o$ is not affected by ρ_{23}. However, $P_{Z^*} Y_o$ is affected (positively in a matrix sense) when Y_o and Y_e are more (positively or negatively) correlated, which explains the increasing probability of detecting endogeneity by the present full-set tests. Finally the two degrees of freedom full-set tests demonstrate power, also when the null hypothesis tested is only partly false. This occurs especially under multicollinearity, which implies when not employing sub-set tests one may unnecessarily declare an exogenous regressor endogenous. Although one would expect lower rejection probability here than for the full-set test which correctly exploits orthogonality of $y^{(3)}$, their comparison is hampered again due to the differences between type I error probabilities. Note though that the first five cases show larger type I error probabilities for T^{23} than for T_3^2, whereas cases 6-10 show fewer correct rejections, which fully conforms to our expectations.

For a higher degree of simultaneity in $y^{(2)}$ (cases 11-13) we find for the sub-set tests that W^3 still underrejects substantially but an effect of multicollinearity is no longer established, which is probably because DGP's with a similar ρ_2 and ρ_3 but higher ρ_{23} are not feasible. Here D^3 does no longer outperform T^3. For the other tests the rejection probabilities that should increase with $|\rho_2|$ do indeed, and we find that the probability of misguidance by the full-set tests exploiting an invalid instrument is even more troublesome now.

These results (which also hold for $n = 100$) already indicate that sub-set tests are indispensable in a comprehensive sequential strategy to classify regressors as either endogenous or exogenous. Because, after a two degrees of freedom full-set test may have indicated that at least one of the two regressors is endogenous, neither of the one degree of freedom full-set tests will be capable of indicating which one is endogenous if there is one endogenous and one exogenous regressor, unless these two regressors are mutually orthogonal. However, the two sub-set tests demonstrate that they can be used to diagnose the endogeneity/exogeneity of the regressors, especially when the endogeneity is serious, irrespective of their degree of multicollinearity. We shall now examine how these capabilities are affected by the strength of the instruments. Because of the poor performance of test H_o we skip it.

The results in Table 2 stem from similar DGP's which differ from the previous ones only in the increased strength of both the instruments, which forces further limitations on multicollinearity, due to (4.6). Note that the size properties have not really improved. Due to the limitations on varying multicollinearity its effects can hardly be assessed from this table. The rejection probabilities of false null hypotheses are larger when the maintained hypothesis is valid, whereas the tests which impose an invalid orthogonality condition become even more confusing when the genuine instruments are stronger. Multicollinearity still has an increasing effect on the rejection probability of all the full-set tests, which is very uncomfortable for the implementations which impose a false exogeneity assumption.

Table 2: One endogenous regressor and stronger instruments:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.30$, $R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
14b	0.0	0.0	0.0	0.049	0.059	0.070	0.046	0.055	0.066	0.050	0.049	0.071	0.049	0.048	0.046	0.069	0.046	0.040	0.043	0.083
14c	0.0	0.0	0.0	0.045	0.053	0.066	0.046	0.055	0.066	0.050	0.047	0.067	0.046	0.049	0.047	0.069	0.046	0.039	0.042	0.085
15b	0.0	0.0	-0.2	0.044	0.055	0.067	0.045	0.056	0.067	0.055	0.050	0.074	0.049	0.051	0.047	0.070	0.046	0.040	0.044	0.084
15c	0.0	0.0	-0.2	0.043	0.054	0.068	0.042	0.054	0.067	0.052	0.048	0.068	0.047	0.053	0.048	0.070	0.047	0.040	0.043	0.085
16b	0.0	0.0	0.2	0.044	0.058	0.068	0.044	0.055	0.069	0.052	0.048	0.072	0.047	0.054	0.051	0.070	0.049	0.040	0.043	0.086
16c	0.0	0.0	0.2	0.042	0.053	0.066	0.043	0.053	0.066	0.049	0.046	0.067	0.045	0.054	0.050	0.072	0.049	0.040	0.043	0.084
17b	0.2	0.0	0.0	0.048	0.059	0.070	0.328	0.357	0.392	0.046	0.045	0.066	0.043	0.329	0.323	0.385	0.322	0.224	0.244	0.351
17c	0.2	0.0	0.0	0.045	0.054	0.066	0.329	0.357	0.392	0.044	0.043	0.065	0.041	0.327	0.323	0.388	0.322	0.222	0.238	0.348
18b	0.2	0.0	-0.2	0.043	0.056	0.066	0.305	0.330	0.373	0.212	0.202	0.252	0.195	0.476	0.461	0.529	0.458	0.356	0.365	0.482
18c	0.2	0.0	-0.2	0.042	0.056	0.068	0.302	0.329	0.375	0.210	0.199	0.254	0.192	0.478	0.464	0.534	0.461	0.357	0.363	0.482
19b	0.2	0.0	0.2	0.045	0.059	0.067	0.305	0.332	0.376	0.218	0.208	0.262	0.202	0.487	0.473	0.539	0.468	0.357	0.365	0.490
19c	0.2	0.0	0.2	0.043	0.054	0.067	0.306	0.331	0.373	0.216	0.204	0.260	0.198	0.486	0.471	0.540	0.467	0.357	0.364	0.489
20b	0.5	0.0	0.0	0.043	0.063	0.067	0.999	0.999	1.000	0.023	0.022	0.035	0.020	1.000	0.999	1.000	0.999	0.998	0.998	0.999
20c	0.5	0.0	0.0	0.041	0.060	0.063	0.999	1.000	1.000	0.024	0.023	0.037	0.022	0.999	0.999	1.000	0.999	0.999	0.998	1.000
21b	0.5	0.0	-0.2	0.040	0.059	0.063	0.993	0.991	0.996	0.980	0.977	0.987	0.957	1.000	1.000	1.000	1.000	1.000	1.000	1.000
21c	0.5	0.0	-0.2	0.043	0.062	0.065	0.993	0.991	0.996	0.981	0.978	0.987	0.955	1.000	1.000	1.000	1.000	1.000	1.000	1.000
22b	0.5	0.0	0.2	0.045	0.063	0.067	0.993	0.991	0.996	0.979	0.976	0.987	0.954	1.000	1.000	1.000	1.000	1.000	1.000	1.000
22c	0.5	0.0	0.2	0.040	0.058	0.063	0.993	0.992	0.995	0.980	0.977	0.987	0.955	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Staiger and Stock (1997) found that full-set tests have correct asymptotic size, although being inconsistent under weak instrument asymptotics. The following three tables illustrate cases in which the instruments are weak for one of the two potentially endogenous variables or for both.

In the DGP's used to generate Table 3, the instruments are weak for $y^{(2)}$ but strong for $y^{(3)}$. So now the two implementations of the sub-set tests examine different situations (even when $\rho_2 = \rho_3 = 0$) and so do the two one degree of freedom full-set test implementations. Especially the sub-set W_o tests and the two degrees of freedom W^{23} test are seriously undersized. When the endogeneity of the weakly instrumented regressor is tested by W_3^2 the type I error probability is seriously affected by (lack of) multicollinearity. All full-set T_o tests are oversized. Only the D_o tests would require just a (mostly) moderate size correction. When $\rho_{23} = 0$ test S_3^2 is oversized. The probability that sub-set test $D^2 (= S^2)$ will detect the endogeneity is small, which was already predicted immediately below (2.18). D_2^3 and S_2^3 will again provide confusing evidence, unless the regressors are orthogonal. Full set tests D_3^2, S_3^2 and $D^{23} (= S^{23})$ have power only under multicollinearity. The latter result can be understood upon specializing (2.18) for this case, where the contributions with c_1 and c_3 disappear because $K_e = 0$ and $L_1 = 0$. Using $\Sigma_{Z'Z} = I$ and $\Sigma_{Y_o'Y_o} = I$ we have to find a solution c_2 satisfying $c_2 = \sigma^2 \gamma_o \neq 0$ and $\Pi_o c_2 = 0$. Since $\gamma_o = (\rho_2 \ 0)'$ and the first column of Π_o vanishes asymptotically there is such a solution indeed, but not if $\Sigma_{Y_o'Y_o}$ were nondiagonal.

The situation is reversed in Table 4, where the instruments are weak for $y^{(3)}$ and strong for the possibly endogenous $y^{(2)}$. Cases 23 and 24 are mirrored in cases 29

Table 3: One endogenous regressor and weak instruments for $y^{(2)}$:

$R^2_{2;z2} = 0.01$, $R^2_{2;z23} = 0.02$, $R^2_{3;z2} = 0.30$, $R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
23b	0.0	0.0	0.0	0.012	0.034	0.022	0.001	0.059	0.055	0.050	0.048	0.073	0.047	0.002	0.044	0.070	0.073	0.005	0.044	0.086
23c	0.0	0.0	0.0	0.012	0.032	0.020	0.001	0.058	0.055	0.048	0.047	0.068	0.046	0.002	0.049	0.073	0.077	0.004	0.040	0.086
24b	0.0	0.0	0.5	0.002	0.058	0.015	0.001	0.059	0.061	0.056	0.049	0.071	0.048	0.045	0.046	0.071	0.047	0.006	0.045	0.085
24c	0.0	0.0	0.5	0.002	0.059	0.014	0.001	0.058	0.062	0.052	0.045	0.069	0.043	0.045	0.046	0.068	0.046	0.006	0.042	0.087
25b	0.2	0.0	0.0	0.012	0.033	0.022	0.001	0.060	0.056	0.048	0.047	0.072	0.046	0.002	0.048	0.069	0.075	0.005	0.047	0.088
25c	0.2	0.0	0.0	0.012	0.034	0.021	0.001	0.061	0.056	0.048	0.047	0.071	0.046	0.003	0.052	0.074	0.081	0.004	0.044	0.088
26b	0.2	0.0	0.5	0.004	0.059	0.029	0.001	0.063	0.064	0.337	0.317	0.381	0.312	0.309	0.315	0.382	0.317	0.090	0.244	0.350
26c	0.2	0.0	0.5	0.003	0.060	0.029	0.001	0.063	0.065	0.338	0.317	0.381	0.312	0.309	0.317	0.378	0.319	0.089	0.241	0.348
27b	0.5	0.0	0.0	0.012	0.039	0.022	0.001	0.083	0.079	0.050	0.047	0.073	0.047	0.002	0.061	0.085	0.094	0.005	0.059	0.112
27c	0.5	0.0	0.0	0.011	0.036	0.020	0.002	0.085	0.079	0.047	0.046	0.069	0.046	0.003	0.064	0.092	0.098	0.005	0.063	0.109
28b	0.5	0.0	0.5	0.026	0.066	0.124	0.001	0.088	0.103	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.699	1.000	1.000
28c	0.5	0.0	0.5	0.026	0.067	0.124	0.002	0.090	0.102	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	0.702	1.000	1.000

Table 4: One endogenous regressor and weak instruments for $y^{(3)}$:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.01$, $R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
29b	0.0	0.0	0.0	0.000	0.054	0.050	0.013	0.033	0.021	0.002	0.047	0.068	0.073	0.048	0.045	0.070	0.045	0.005	0.044	0.082
29c	0.0	0.0	0.0	0.001	0.053	0.051	0.012	0.030	0.021	0.002	0.047	0.070	0.072	0.047	0.045	0.070	0.045	0.006	0.043	0.083
30b	0.0	0.0	0.5	0.001	0.056	0.058	0.001	0.053	0.012	0.046	0.048	0.070	0.048	0.056	0.048	0.071	0.047	0.007	0.043	0.085
30c	0.0	0.0	0.5	0.000	0.056	0.058	0.002	0.054	0.014	0.047	0.049	0.071	0.050	0.056	0.048	0.071	0.047	0.006	0.043	0.085
31b	0.2	0.0	0.0	0.000	0.054	0.055	0.098	0.135	0.139	0.005	0.156	0.196	0.248	0.327	0.322	0.386	0.321	0.070	0.238	0.347
31c	0.2	0.0	0.0	0.001	0.054	0.054	0.104	0.136	0.147	0.006	0.155	0.198	0.249	0.327	0.323	0.388	0.322	0.069	0.237	0.347
32b	0.2	0.0	0.5	0.001	0.056	0.067	0.017	0.072	0.093	0.863	0.865	0.898	0.866	0.887	0.876	0.906	0.872	0.439	0.807	0.880
32c	0.2	0.0	0.5	0.001	0.057	0.061	0.018	0.072	0.091	0.860	0.863	0.897	0.863	0.888	0.875	0.906	0.871	0.435	0.807	0.880
33b	0.5	0.0	0.0	0.001	0.060	0.071	0.566	0.453	0.639	0.065	0.607	0.639	0.818	0.999	0.999	1.000	0.999	0.663	0.998	0.999
33c	0.5	0.0	0.0	0.001	0.059	0.071	0.578	0.459	0.647	0.067	0.602	0.634	0.816	0.999	0.999	1.000	0.999	0.674	0.998	1.000
34b	0.5	0.0	0.2	0.001	0.061	0.076	0.423	0.297	0.546	0.432	0.865	0.881	0.949	1.000	1.000	1.000	1.000	0.709	1.000	1.000
34c	0.5	0.0	0.2	0.001	0.061	0.074	0.427	0.308	0.540	0.432	0.867	0.880	0.945	1.000	1.000	1.000	1.000	0.704	1.000	1.000

and 30. The W_o tests are seriously undersized, except W_3^2 (building on exogeneity of $y^{(3)}$, it is not affected by its weak external instruments) and W_2^3 (provided the multicollinearity is substantial). The full-set T_o tests are again oversized. Some D_o (and more S_o) implementations show some size distortions. Because of the findings below (2.18) it is no surprise that the sub-set tests on $y^{(2)}$ exhibit deminishing power for more severe multicollinearity. After size correction it seems likely that W^2 or especially T^2 would do better than $D^2(= S^2)$. All the tests that exploit the exogeneity of $y^{(3)}$ show power for detecting endogeneity of $y^{(2)}$ when the instruments are weak for exogenous regressor $y^{(3)}$, and their power increases with multicollinearity and of course with ρ_2.

Next we construct DGP's in which the instruments are weak for both regressors. Given our predictions below (2.18) and because we found mixed results when the instruments are weak for one of the two regressors, not much should be expected when both are affected. The results in Table 5 do indeed illustrate this. The W_o tests underreject severely, T_o gives a mixed picture, but D_o would require only a minor (and S_o a more substantial) size correction, although they will yield very modest power.

In addition to cases in which the two instruments have similar strength for $y^{(2)}$

Table 5: One endogenous regressor and weak instruments:

$R^2_{2;z2} = 0.01$, $R^2_{2;z23} = 0.02$, $R^2_{3;z2} = 0.01$, $R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
35b	0.0	0.0	0.0	0.000	0.032	0.016	0.001	0.031	0.017	0.001	0.051	0.074	0.077	0.002	0.048	0.070	0.076	0.000	0.042	0.086
35c	0.0	0.0	0.0	0.000	0.030	0.018	0.000	0.034	0.016	0.002	0.048	0.070	0.076	0.002	0.049	0.072	0.078	0.000	0.041	0.086
36b	0.0	0.0	0.5	0.000	0.036	0.018	0.000	0.036	0.020	0.002	0.048	0.069	0.071	0.002	0.044	0.067	0.071	0.000	0.043	0.083
36c	0.0	0.0	0.5	0.000	0.034	0.016	0.000	0.037	0.019	0.003	0.047	0.067	0.069	0.003	0.049	0.072	0.074	0.000	0.041	0.087
37b	0.2	0.0	0.0	0.000	0.033	0.018	0.000	0.032	0.018	0.001	0.050	0.073	0.080	0.003	0.050	0.072	0.078	0.000	0.046	0.088
37c	0.2	0.0	0.0	0.000	0.033	0.016	0.000	0.035	0.018	0.002	0.048	0.070	0.076	0.003	0.051	0.075	0.078	0.000	0.045	0.088
38b	0.2	0.0	0.5	0.000	0.038	0.021	0.000	0.039	0.020	0.002	0.051	0.076	0.075	0.003	0.049	0.072	0.074	0.000	0.047	0.092
38c	0.2	0.0	0.5	0.000	0.033	0.018	0.000	0.039	0.022	0.002	0.051	0.073	0.077	0.002	0.052	0.078	0.080	0.000	0.048	0.092
39b	0.5	0.0	0.0	0.000	0.037	0.019	0.000	0.040	0.026	0.002	0.057	0.086	0.092	0.002	0.063	0.087	0.095	0.000	0.058	0.107
39c	0.5	0.0	0.0	0.000	0.036	0.020	0.000	0.041	0.022	0.002	0.056	0.081	0.090	0.003	0.064	0.090	0.097	0.000	0.062	0.108
40b	0.5	0.0	0.5	0.000	0.048	0.025	0.000	0.051	0.030	0.005	0.083	0.116	0.123	0.007	0.089	0.122	0.130	0.000	0.087	0.148
40c	0.5	0.0	0.5	0.000	0.049	0.025	0.001	0.057	0.030	0.004	0.081	0.112	0.118	0.006	0.092	0.125	0.132	0.000	0.089	0.151

and $y^{(3)}$, we present a couple of cases in which this differs. Note that the inequality (4.7) is now satisfied by all four combinations in (4.25). The reason that not every case in Table 6 consists of four subcases is that not every subcase satisfies the second part of (4.6). The results for the sub-set tests can differ greatly between the four subcases. Subcases "a" and "d" show lower rejection probabilities for W_o and T_o, whereas D_o and S_o seem hardly affected under the null hypothesis. This suggests that the estimate $\hat{\beta}_r$ (and hence $\hat{\sigma}_r^2$) is probably less affected by (d_{23}, d_{33}) in these subcases than $\hat{\sigma}^2$ and $\ddot{\sigma}^2$.

The sub-set tests on $y^{(2)}$ and $y^{(3)}$ behave similarly although the (joint) instrument strength is a little higher for the former. Whereas the results between the subcases are quite different for the sub-set tests and the two degrees of freedom full-set tests, the one degree of freedom full-set test seem less dependent on the choice of (d_{23}, d_{33}).

When $y^{(2)}$ is endogenous D^2 has substantially less power in subcases "a" and "d" even though under the null hypothesis it rejects less often in subcases "c" and "d". For the full-set test things are different. These reject far more often in subcases "a" and "d" when there is little or no multicollinearity. However, when multicollinearity is more pronounced the tests reject less often in subcases "a" and "d" than in "b" and "c". From these results we conclude that the relevant nuisance parameters for these asymptotic tests are not just simultaneity, multicollinearity and instrument strength, but also the actual signs of the reduced form coefficients.

5.2 Both regressors endogenous

The rejection probabilities of the sub-set tests estimated under the alternative hypothesis in the previous subsection are only of secondary interest, because the sub-set that was not tested and allowed to be endogenous was actually exogenous. In such cases application of the one-degree of freedom full-set test is more appropriate. Now the not tested sub-set which is treated as endogenous will actually be endogenous, so we will get crucial information on the practical usefulness of the sub-set tests, and further evidence on the possible misguidance by the here inappropriate one degree of

Table 6: One endogenous regressor and asymmetric instrument strength:

$R^2_{2;z2} = 0.30,\ R^2_{2;z23} = 0.50,\ R^2_{3;z2} = 0.10,\ R^2_{3;z23} = 0.40$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
41a	0.0	0.0	0.0	0.002	0.054	0.030	0.001	0.054	0.026	0.053	0.048	0.070	0.047	0.056	0.049	0.070	0.047	0.007	0.042	0.085
41b	0.0	0.0	0.0	0.032	0.054	0.066	0.038	0.054	0.063	0.041	0.050	0.073	0.051	0.044	0.048	0.070	0.048	0.024	0.042	0.083
41c	0.0	0.0	0.0	0.033	0.052	0.064	0.035	0.050	0.063	0.038	0.046	0.066	0.047	0.046	0.048	0.069	0.048	0.026	0.044	0.083
41d	0.0	0.0	0.0	0.001	0.056	0.035	0.002	0.056	0.027	0.054	0.049	0.072	0.048	0.056	0.050	0.071	0.048	0.007	0.041	0.084
42b	0.0	0.0	-0.2	0.026	0.052	0.063	0.030	0.052	0.060	0.047	0.048	0.072	0.048	0.050	0.048	0.072	0.048	0.026	0.042	0.082
42c	0.0	0.0	-0.2	0.034	0.053	0.064	0.040	0.051	0.063	0.039	0.046	0.067	0.048	0.045	0.047	0.069	0.047	0.027	0.042	0.085
42d	0.0	0.0	-0.2	0.001	0.056	0.024	0.002	0.054	0.018	0.045	0.049	0.073	0.049	0.048	0.050	0.072	0.050	0.007	0.043	0.085
43a	0.0	0.0	0.2	0.002	0.052	0.023	0.002	0.051	0.017	0.043	0.049	0.070	0.049	0.048	0.048	0.072	0.048	0.007	0.042	0.085
43b	0.0	0.0	0.2	0.034	0.055	0.065	0.039	0.053	0.064	0.038	0.047	0.070	0.049	0.044	0.046	0.070	0.047	0.027	0.042	0.085
43c	0.0	0.0	0.2	0.029	0.052	0.061	0.027	0.050	0.059	0.046	0.046	0.068	0.047	0.051	0.049	0.071	0.049	0.026	0.043	0.084
44c	0.0	0.0	-0.5	0.032	0.058	0.068	0.032	0.055	0.064	0.052	0.047	0.069	0.046	0.054	0.049	0.069	0.048	0.029	0.041	0.084
44d	0.0	0.0	-0.5	0.004	0.051	0.025	0.008	0.047	0.021	0.021	0.048	0.069	0.054	0.034	0.047	0.068	0.050	0.008	0.043	0.084
45a	0.0	0.0	0.5	0.003	0.051	0.024	0.007	0.046	0.022	0.021	0.048	0.070	0.054	0.033	0.046	0.068	0.049	0.007	0.042	0.082
45b	0.0	0.0	0.5	0.033	0.060	0.071	0.034	0.056	0.067	0.053	0.049	0.070	0.048	0.052	0.047	0.070	0.046	0.030	0.044	0.087
46a	0.2	0.0	0.0	0.002	0.055	0.044	0.006	0.067	0.070	0.697	0.679	0.737	0.675	0.711	0.691	0.750	0.685	0.300	0.586	0.696
46b	0.2	0.0	0.0	0.032	0.055	0.066	0.186	0.230	0.262	0.054	0.062	0.088	0.066	0.222	0.232	0.287	0.233	0.120	0.169	0.260
46c	0.2	0.0	0.0	0.032	0.053	0.064	0.190	0.234	0.266	0.052	0.061	0.088	0.064	0.222	0.231	0.290	0.233	0.123	0.170	0.262
46d	0.2	0.0	0.0	0.002	0.056	0.044	0.006	0.069	0.072	0.708	0.693	0.748	0.686	0.719	0.700	0.757	0.694	0.310	0.595	0.706
47b	0.2	0.0	-0.2	0.027	0.053	0.062	0.154	0.198	0.238	0.185	0.191	0.239	0.191	0.347	0.343	0.409	0.340	0.210	0.259	0.365
47c	0.2	0.0	-0.2	0.033	0.054	0.064	0.201	0.242	0.273	0.050	0.058	0.085	0.063	0.230	0.239	0.294	0.240	0.128	0.177	0.266
47d	0.2	0.0	-0.2	0.003	0.056	0.031	0.010	0.075	0.054	0.245	0.262	0.326	0.267	0.289	0.290	0.350	0.290	0.082	0.214	0.314
48a	0.2	0.0	0.2	0.003	0.053	0.029	0.012	0.073	0.054	0.249	0.264	0.327	0.267	0.282	0.286	0.348	0.286	0.081	0.216	0.312
48b	0.2	0.0	0.2	0.033	0.056	0.066	0.203	0.243	0.273	0.051	0.061	0.089	0.065	0.230	0.238	0.298	0.240	0.126	0.179	0.273
48c	0.2	0.0	0.2	0.028	0.053	0.060	0.158	0.200	0.241	0.186	0.188	0.241	0.190	0.347	0.343	0.411	0.340	0.210	0.258	0.361
49c	0.2	0.0	-0.5	0.032	0.058	0.067	0.171	0.203	0.252	0.616	0.599	0.663	0.591	0.725	0.709	0.761	0.705	0.587	0.604	0.715
49d	0.2	0.0	-0.5	0.007	0.052	0.030	0.033	0.102	0.077	0.054	0.112	0.153	0.130	0.136	0.169	0.219	0.176	0.041	0.130	0.207
50a	0.2	0.0	0.5	0.006	0.054	0.029	0.033	0.101	0.072	0.055	0.111	0.152	0.127	0.128	0.166	0.215	0.173	0.038	0.123	0.202
50b	0.2	0.0	0.5	0.032	0.060	0.070	0.168	0.203	0.253	0.611	0.597	0.665	0.590	0.725	0.709	0.767	0.705	0.590	0.607	0.716
51b	0.5	0.0	0.0	0.029	0.061	0.064	0.947	0.952	0.967	0.148	0.169	0.218	0.206	0.975	0.975	0.984	0.975	0.937	0.957	0.976
51c	0.5	0.0	0.0	0.028	0.059	0.060	0.947	0.951	0.967	0.147	0.167	0.213	0.203	0.976	0.976	0.984	0.975	0.937	0.957	0.979
52b	0.5	0.0	-0.2	0.027	0.059	0.060	0.842	0.836	0.903	0.933	0.933	0.954	0.932	1.000	1.000	1.000	1.000	0.996	1.000	1.000
52c	0.5	0.0	-0.2	0.028	0.059	0.060	0.960	0.963	0.973	0.144	0.164	0.207	0.201	0.984	0.984	0.990	0.985	0.958	0.969	0.985
52d	0.5	0.0	-0.2	0.014	0.061	0.086	0.101	0.193	0.326	0.990	0.991	0.995	0.992	0.998	0.998	0.999	0.998	0.720	0.996	0.998
53a	0.5	0.0	0.2	0.012	0.059	0.086	0.097	0.189	0.317	0.990	0.992	0.995	0.992	0.997	0.997	0.998	0.996	0.699	0.994	0.997
53b	0.5	0.0	0.2	0.030	0.061	0.063	0.959	0.962	0.972	0.149	0.169	0.213	0.205	0.984	0.983	0.988	0.983	0.955	0.968	0.985
53c	0.5	0.0	0.2	0.027	0.057	0.060	0.846	0.842	0.910	0.933	0.934	0.954	0.932	1.000	1.000	1.000	1.000	0.996	1.000	1.000
54d	0.5	0.0	-0.5	0.024	0.063	0.073	0.310	0.423	0.510	0.437	0.619	0.674	0.699	0.868	0.905	0.930	0.911	0.503	0.869	0.922
55a	0.5	0.0	0.5	0.023	0.063	0.077	0.298	0.418	0.499	0.441	0.623	0.675	0.698	0.864	0.905	0.931	0.911	0.489	0.869	0.921

freedom full-set tests. Similar cases in terms of instrument strength have been chosen to keep comparability with the previous subsection.

The DGP's used for Table 7 mimic those of Table 1 in terms of instrument strength. In most cases the sub-set tests behave roughly the same as when the maintained regressor was actually exogenous, although multicollinearity is now found to have a small though clear asymmetric impact on the rejection probabilities. When multicollinearity is of the same sign as the simultaneity in $y^{(3)}$, test statistics W_o and T_o reject less often than when these signs differ. This is not caused by the fixed nature of the instruments, because simulations (not reported) in which the instruments are random show the same effect. On the other hand, the differences between subcases diminish when the instruments are random. Multicollinearity decreases the rejection probabilities, but less so when the endogeneity of the maintained regressor is more severe. The full-set tests with one degree of freedom are affected more by multi-

Table 7: Two endogenous regressors and mildly strong instruments:

$R^2_{2;z2} = 0.20$, $R^2_{2;z23} = 0.40$, $R^2_{3;z2} = 0.20$, $R^2_{3;z23} = 0.40$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W^3_2	D^3_2	T^3_2	S^3_2	W^2_3	D^2_3	T^2_3	S^2_3	W^{23}	D^{23}	T^{23}
57b	0.2	-0.2	0.0	0.126	0.183	0.203	0.125	0.183	0.202	0.146	0.173	0.217	0.180	0.141	0.166	0.211	0.174	0.129	0.222	0.329
57c	0.2	-0.2	0.0	0.124	0.181	0.200	0.124	0.179	0.199	0.142	0.166	0.213	0.173	0.141	0.165	0.211	0.173	0.128	0.223	0.325
58b	0.2	-0.2	-0.2	0.109	0.175	0.191	0.103	0.171	0.184	0.076	0.085	0.117	0.087	0.073	0.084	0.114	0.087	0.070	0.147	0.230
58c	0.2	-0.2	-0.2	0.105	0.173	0.188	0.107	0.176	0.188	0.073	0.083	0.111	0.085	0.072	0.081	0.112	0.084	0.072	0.146	0.230
59b	0.2	-0.2	0.2	0.131	0.173	0.208	0.132	0.175	0.209	0.381	0.403	0.468	0.412	0.385	0.408	0.477	0.415	0.334	0.434	0.552
59c	0.2	-0.2	0.2	0.128	0.171	0.207	0.130	0.174	0.208	0.382	0.406	0.477	0.414	0.382	0.406	0.484	0.415	0.331	0.427	0.551
60b	0.2	-0.2	-0.5	0.071	0.148	0.166	0.071	0.146	0.163	0.054	0.047	0.074	0.045	0.053	0.048	0.069	0.046	0.046	0.097	0.165
60c	0.2	-0.2	-0.5	0.073	0.150	0.165	0.072	0.154	0.170	0.056	0.049	0.073	0.047	0.054	0.047	0.069	0.045	0.046	0.101	0.167
61b	0.2	-0.2	0.5	0.119	0.148	0.201	0.120	0.152	0.207	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
61c	0.2	-0.2	0.5	0.117	0.146	0.200	0.117	0.146	0.206	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
62b	0.2	0.2	0.0	0.122	0.177	0.199	0.124	0.182	0.201	0.140	0.164	0.208	0.172	0.140	0.165	0.212	0.171	0.128	0.221	0.324
62c	0.2	0.2	0.0	0.124	0.177	0.197	0.126	0.179	0.199	0.141	0.164	0.208	0.171	0.143	0.165	0.211	0.172	0.129	0.222	0.326
63b	0.2	0.2	-0.2	0.127	0.170	0.205	0.125	0.169	0.205	0.373	0.400	0.469	0.408	0.381	0.406	0.473	0.415	0.328	0.424	0.538
63c	0.2	0.2	-0.2	0.129	0.172	0.205	0.133	0.175	0.207	0.381	0.407	0.478	0.417	0.378	0.403	0.476	0.412	0.327	0.424	0.544
64b	0.2	0.2	0.2	0.107	0.174	0.185	0.106	0.176	0.189	0.070	0.081	0.108	0.084	0.074	0.085	0.117	0.089	0.067	0.147	0.231
64c	0.2	0.2	0.2	0.104	0.168	0.182	0.108	0.172	0.186	0.072	0.079	0.110	0.082	0.075	0.085	0.117	0.089	0.069	0.147	0.229
65b	0.2	0.2	-0.5	0.113	0.145	0.205	0.115	0.146	0.204	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000
65c	0.2	0.2	-0.5	0.115	0.145	0.201	0.120	0.151	0.206	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
66b	0.2	0.2	0.5	0.073	0.150	0.166	0.074	0.151	0.171	0.055	0.049	0.070	0.046	0.053	0.048	0.071	0.046	0.043	0.098	0.168
66c	0.2	0.2	0.5	0.069	0.147	0.162	0.072	0.148	0.166	0.051	0.045	0.070	0.044	0.053	0.047	0.072	0.045	0.042	0.096	0.165
67b	0.5	-0.2	0.0	0.127	0.201	0.208	0.814	0.861	0.882	0.182	0.208	0.258	0.240	0.848	0.867	0.901	0.875	0.771	0.886	0.930
67c	0.5	-0.2	0.0	0.124	0.199	0.206	0.816	0.863	0.884	0.182	0.209	0.262	0.244	0.855	0.874	0.906	0.882	0.775	0.883	0.935
68b	0.5	-0.2	-0.2	0.093	0.193	0.177	0.780	0.852	0.874	0.049	0.056	0.080	0.068	0.737	0.754	0.805	0.760	0.592	0.770	0.848
68c	0.5	-0.2	-0.2	0.089	0.192	0.173	0.779	0.851	0.875	0.048	0.054	0.079	0.064	0.736	0.754	0.806	0.762	0.591	0.766	0.846
69b	0.5	-0.2	0.2	0.151	0.190	0.226	0.754	0.779	0.840	0.876	0.886	0.917	0.899	0.997	0.998	0.998	0.997	0.992	0.999	1.000
69c	0.5	-0.2	0.2	0.150	0.186	0.224	0.757	0.776	0.837	0.884	0.893	0.922	0.905	0.997	0.997	0.999	0.997	0.993	0.999	1.000
70b	0.5	-0.2	-0.5	0.040	0.166	0.134	0.595	0.690	0.767	0.996	0.995	0.997	0.992	1.000	1.000	1.000	1.000	0.989	1.000	1.000
70c	0.5	-0.2	-0.5	0.043	0.163	0.133	0.595	0.691	0.767	0.995	0.994	0.997	0.990	1.000	1.000	1.000	1.000	0.989	1.000	1.000
71b	0.5	0.2	0.0	0.124	0.195	0.201	0.815	0.861	0.882	0.180	0.206	0.256	0.240	0.848	0.866	0.899	0.874	0.772	0.883	0.932
71c	0.5	0.2	0.0	0.126	0.196	0.202	0.810	0.859	0.886	0.176	0.202	0.253	0.236	0.846	0.866	0.900	0.875	0.772	0.881	0.931
72b	0.5	0.2	-0.2	0.147	0.186	0.224	0.752	0.773	0.835	0.884	0.894	0.922	0.907	0.997	0.997	0.999	0.998	0.992	0.999	1.000
72c	0.5	0.2	-0.2	0.149	0.184	0.222	0.754	0.772	0.838	0.880	0.890	0.919	0.902	0.997	0.997	0.998	0.997	0.993	0.998	0.999
73b	0.5	0.2	0.2	0.090	0.193	0.170	0.783	0.854	0.872	0.050	0.057	0.083	0.067	0.738	0.756	0.809	0.763	0.597	0.768	0.850
73c	0.5	0.2	0.2	0.087	0.189	0.168	0.787	0.855	0.875	0.047	0.055	0.078	0.065	0.741	0.757	0.809	0.764	0.595	0.769	0.845
74b	0.5	0.2	0.5	0.045	0.166	0.138	0.598	0.692	0.767	0.997	0.995	0.998	0.992	1.000	1.000	1.000	1.000	0.988	1.000	1.000
74c	0.5	0.2	0.5	0.044	0.162	0.132	0.596	0.688	0.767	0.996	0.995	0.997	0.990	1.000	1.000	1.000	1.000	0.988	1.000	1.000
75b	0.5	-0.5	0.0	0.799	0.834	0.866	0.804	0.838	0.870	0.935	0.945	0.962	0.958	0.937	0.949	0.963	0.961	0.985	1.000	1.000
75c	0.5	-0.5	0.0	0.806	0.840	0.872	0.804	0.837	0.865	0.937	0.947	0.962	0.959	0.938	0.947	0.961	0.959	0.987	1.000	1.000
76b	0.5	-0.5	-0.2	0.796	0.891	0.891	0.799	0.886	0.889	0.409	0.433	0.500	0.465	0.412	0.434	0.502	0.465	0.843	0.970	0.985
76c	0.5	-0.5	-0.2	0.804	0.892	0.893	0.803	0.887	0.889	0.412	0.436	0.509	0.469	0.411	0.436	0.498	0.468	0.848	0.971	0.986
77b	0.5	-0.5	-0.5	0.600	0.797	0.810	0.600	0.797	0.812	0.060	0.054	0.079	0.046	0.058	0.052	0.075	0.045	0.348	0.684	0.782
77c	0.5	-0.5	-0.5	0.607	0.803	0.818	0.602	0.799	0.812	0.061	0.055	0.079	0.046	0.056	0.051	0.074	0.043	0.351	0.684	0.784
78b	0.5	0.5	0.0	0.804	0.840	0.869	0.801	0.835	0.867	0.940	0.950	0.965	0.963	0.939	0.950	0.964	0.963	0.985	1.000	1.000
78c	0.5	0.5	0.0	0.802	0.838	0.867	0.801	0.835	0.867	0.938	0.949	0.962	0.961	0.937	0.946	0.963	0.958	0.985	1.000	1.000
79b	0.5	0.5	0.2	0.799	0.891	0.891	0.800	0.886	0.887	0.409	0.431	0.504	0.467	0.413	0.435	0.503	0.468	0.848	0.973	0.987
79c	0.5	0.5	0.2	0.799	0.887	0.888	0.798	0.887	0.888	0.412	0.434	0.503	0.467	0.406	0.429	0.494	0.458	0.847	0.970	0.985
80b	0.5	0.5	0.5	0.598	0.798	0.812	0.606	0.800	0.812	0.052	0.046	0.070	0.039	0.061	0.054	0.079	0.046	0.340	0.680	0.780
80c	0.5	0.5	0.5	0.602	0.797	0.812	0.602	0.800	0.815	0.058	0.052	0.075	0.045	0.061	0.055	0.079	0.047	0.348	0.682	0.782

collinearity than the sub-set tests. As is to be expected, the two degrees of freedom full-set tests reject more often now that both regressors are endogenous. The rejection probabilities of these full-set tests, D_o included, decrease dramatically if ρ_{23} and ρ_3 are of the same sign, and they do that much more than for the sub-set tests. Note that the cases in which ρ_3 takes on a negative value are very similar to cases in which

ρ_3 is positive and the sign of ρ_{23} is changed, or those of (d_{32}, d_{33}). More specifically, case 63b corresponds with case 59c and case 63c with case 59b. Therefore, we will exclude cases with negative values for ρ_3 from future tables and stick to their positive counterparts.

In Table 8 we examine stronger instruments. Comparing with Table 2 we find that the rejection probabilities seem virtually unaffected by choosing $\rho_3 \neq 0$. As we found before the rejection probabilities are affected in a positive manner by the increased strength of the instruments. The sub-set tests reject almost every time if the corresponding degree of simultaneity is .5. The effect of having ρ_{23} and ρ_3 both positive seems less severe. As long as this is not the case, the one degree of freedom full-set tests reject more often than the sub-set tests. If ρ_{23} and ρ_3 do not differ in sign W_o and D_o reject more often when applied to a sub-set than for their one degree and two degrees of freedom full-set versions.

Table 8: Two endogenous regressors and stronger instruments:

$R^2_{2;z2} = 0.30,\ R^2_{2;z23} = 0.60,\ R^2_{3;z2} = 0.30,\ R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
81b	0.2	0.2	0.0	0.322	0.349	0.383	0.330	0.359	0.393	0.327	0.321	0.388	0.319	0.331	0.327	0.394	0.322	0.475	0.494	0.607
81c	0.2	0.2	0.0	0.327	0.357	0.388	0.330	0.361	0.395	0.331	0.326	0.392	0.323	0.330	0.325	0.392	0.323	0.477	0.497	0.620
82b	0.2	0.2	-0.2	0.315	0.321	0.377	0.319	0.327	0.381	0.934	0.929	0.951	0.923	0.934	0.928	0.949	0.921	0.954	0.950	0.973
82c	0.2	0.2	-0.2	0.315	0.323	0.380	0.316	0.324	0.382	0.934	0.926	0.948	0.920	0.931	0.925	0.948	0.920	0.949	0.947	0.972
83b	0.2	0.2	0.2	0.285	0.331	0.360	0.293	0.340	0.368	0.099	0.091	0.125	0.087	0.103	0.097	0.132	0.093	0.237	0.279	0.384
83c	0.2	0.2	0.2	0.291	0.338	0.362	0.295	0.342	0.366	0.098	0.092	0.127	0.087	0.101	0.093	0.131	0.090	0.237	0.275	0.385
84b	0.5	0.2	0.0	0.337	0.370	0.397	0.999	0.999	0.999	0.400	0.394	0.474	0.369	1.000	1.000	1.000	0.999	1.000	1.000	1.000
84c	0.5	0.2	0.0	0.334	0.368	0.397	0.999	0.999	0.999	0.402	0.394	0.472	0.366	1.000	1.000	1.000	0.999	1.000	1.000	1.000
85b	0.5	0.2	0.2	0.273	0.356	0.352	0.998	0.998	0.999	0.194	0.183	0.241	0.148	0.997	0.996	0.998	0.996	0.998	0.998	0.999
85c	0.5	0.2	0.2	0.277	0.363	0.357	0.998	0.998	0.999	0.191	0.181	0.236	0.149	0.997	0.997	0.998	0.996	0.999	0.999	0.999
86b	0.5	0.5	0.2	1.000	1.000	1.000	0.999	0.999	1.000	0.634	0.616	0.693	0.535	0.636	0.617	0.698	0.532	1.000	1.000	1.000
86c	0.5	0.5	0.2	1.000	1.000	1.000	1.000	1.000	1.000	0.637	0.616	0.698	0.533	0.626	0.609	0.690	0.525	1.000	1.000	1.000

Because Tables 3 and 4 are very similar, we only need to consider the equivalent table of the latter when both regressors are endogenous. In Table 9 the instruments are weak for $y^{(3)}$ but strong for $y^{(2)}$. Obviously the sub-set tests for $y^{(3)}$ lack power now, as was already concluded from Table 3. However, sub-set tests for $y^{(2)}$ show power also in the presence of a maintained endogenous though weakly instrumented regressor. Note that when ρ_3 is increased all sub-set tests for $y^{(2)}$ reject more often. This dependence was not apparent under non-weak instruments.

As we found in Table 5 the sub-set tests perform badly when the instruments are weak for both regressors. From the results on the sub-set test for $y^{(3)}$ we expect the same for the case in which $\rho_3 \neq 0$. This we found to be true in further simulations (not reported here).

Tables 7, 8 and 9 demonstrate that the sub-set tests are indispensable when there is more than one regressor that might be endogenous. Using only full-set tests seriously hampers to correctly classify the individual variables as either endogenous or exogenous. However, all tests examined here, especially W_o and T_o, show substantial size distortions in finite samples. Moreover, these size distortions are found to be determined in a complex way by the model characteristics. In fact the various tables

Table 9: Two endogenous regressors and weak instruments for $y^{(3)}$:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.01$, $R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
87b	0.2	0.2	0.0	0.001	0.059	0.057	0.103	0.140	0.147	0.006	0.162	0.206	0.260	0.339	0.334	0.400	0.332	0.074	0.253	0.361
87c	0.2	0.2	0.0	0.001	0.059	0.059	0.109	0.143	0.152	0.006	0.164	0.205	0.260	0.337	0.334	0.400	0.333	0.077	0.251	0.361
88b	0.2	0.2	-0.2	0.001	0.058	0.058	0.099	0.131	0.152	0.095	0.386	0.447	0.467	0.541	0.532	0.605	0.530	0.159	0.435	0.552
88c	0.2	0.2	-0.2	0.001	0.061	0.063	0.103	0.131	0.161	0.096	0.385	0.446	0.469	0.542	0.533	0.604	0.531	0.162	0.434	0.555
89b	0.2	0.2	0.2	0.001	0.059	0.059	0.052	0.103	0.084	0.028	0.164	0.208	0.212	0.259	0.252	0.309	0.251	0.054	0.186	0.284
89c	0.2	0.2	0.2	0.001	0.059	0.056	0.051	0.099	0.088	0.029	0.162	0.206	0.210	0.259	0.252	0.307	0.251	0.054	0.186	0.279
90b	0.2	0.2	-0.5	0.001	0.059	0.075	0.041	0.076	0.191	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.739	1.000	1.000
90c	0.2	0.2	-0.5	0.001	0.060	0.078	0.043	0.075	0.189	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.747	1.000	1.000
91b	0.2	0.2	0.5	0.000	0.060	0.063	0.006	0.069	0.041	0.270	0.277	0.338	0.279	0.314	0.294	0.355	0.289	0.073	0.217	0.318
91c	0.2	0.2	0.5	0.001	0.060	0.063	0.006	0.069	0.040	0.269	0.277	0.336	0.278	0.316	0.293	0.356	0.288	0.076	0.219	0.317
92b	0.5	0.2	0.0	0.001	0.062	0.073	0.578	0.456	0.645	0.070	0.617	0.650	0.828	1.000	1.000	1.000	1.000	0.671	0.999	1.000
92c	0.5	0.2	0.0	0.001	0.065	0.077	0.591	0.470	0.655	0.071	0.611	0.639	0.825	1.000	1.000	1.000	1.000	0.684	0.999	1.000
93b	0.5	0.2	-0.2	0.001	0.064	0.086	0.448	0.295	0.569	0.489	0.905	0.915	0.970	1.000	1.000	1.000	1.000	0.739	1.000	1.000
93c	0.5	0.2	-0.2	0.001	0.065	0.087	0.454	0.300	0.582	0.490	0.902	0.912	0.966	1.000	1.000	1.000	1.000	0.754	1.000	1.000
94b	0.5	0.2	0.2	0.001	0.064	0.077	0.409	0.314	0.529	0.385	0.828	0.848	0.922	0.999	0.999	1.000	0.999	0.666	0.998	0.999
94c	0.5	0.2	0.2	0.001	0.064	0.074	0.415	0.324	0.531	0.382	0.820	0.839	0.919	0.999	0.999	0.999	0.999	0.675	0.998	0.999
95b	0.5	0.5	0.0	0.002	0.093	0.110	0.608	0.475	0.676	0.096	0.672	0.700	0.871	1.000	1.000	1.000	1.000	0.705	1.000	1.000
95c	0.5	0.5	0.0	0.001	0.092	0.112	0.615	0.486	0.680	0.101	0.676	0.701	0.868	1.000	1.000	1.000	1.000	0.705	1.000	1.000
96b	0.5	0.5	0.2	0.002	0.088	0.102	0.418	0.359	0.541	0.342	0.791	0.813	0.900	0.999	0.999	0.999	0.998	0.629	0.998	0.999
96c	0.5	0.5	0.2	0.002	0.090	0.102	0.426	0.376	0.551	0.347	0.780	0.804	0.894	0.999	0.998	0.999	0.998	0.637	0.997	0.999
97b	0.5	0.5	0.5	0.001	0.088	0.103	0.067	0.159	0.285	0.998	0.998	0.999	0.998	1.000	1.000	1.000	1.000	0.674	0.998	1.000
97c	0.5	0.5	0.5	0.001	0.088	0.101	0.070	0.161	0.282	0.998	0.998	0.999	0.998	1.000	1.000	1.000	0.999	0.677	0.999	1.000

illustrate that it are not just the design parameters simultaneity, multicollinearity and instrument strength which determine the size of these tests. The differences between the subcases illustrate that the size also depends on the actual reduced form coefficients and therefore in fact on the degree by which the multicollinearity stems from correlation between the reduced form disturbances (κ). Trying to mitigate the size problems by simple degrees of freedom adjustments or by transformations to F statistics seems therefore a dead-end.

6 Results for bootstrapped tests

Because all the test statistics that are under investigation here are based on appropriate first order asymptotics, it should be feasible to mitigate the size problems by bootstrapping.

6.1 A bootstrap routine for sub-set DWH test statistics

Bootstrap routines for testing the orthogonality of all possibly endogenous regressors have previously been discussed by Wong (1996). Implementation of these bootstrap routines is relatively easy due to the fact that no regressors are assumed to be endogenous under the null hypothesis. This in contrast to the test of sub-sets where some regressors are endogenous also under the null hypothesis. Their presence complicates matters as bootstrap realizations have to be generated on both the dependent variable and the maintained set of endogenous regressors. We discuss two routines; first a parametric and next a semiparametric bootstrap. For the former routine we have to assume a distribution for the disturbances, which we choose to be the normal.

Consider the $n \times (1+K_e)$ matrix $U = (u, V_e)$. Its elements can be estimated by: $\hat{u}_r = y - X\hat{\beta}_r$ and $\hat{V}_{er} = Y_e - Z_r\hat{\Pi}_{er}$, where $\hat{\Pi}_{er} = (Z_r'Z_r)^{-1}Z_r'Y_e$. Under the null hypothesis $\hat{\beta}_r$ and $\hat{\Pi}_{er}$ are consistent estimators and it follows that $\hat{U}_r = (\hat{u}_r, \hat{V}_{er})$ is consistent for U, and hence $\hat{\Sigma} = n^{-1}\hat{U}_r'\hat{U}_r$ is a consistent estimator of the variance of its rows. The following illustrates the steps that are required for the bootstrap procedure.

1. Draw pseudo disturbances of sample size n from the $N(0, \hat{\Sigma})$ distribution and collect them in $U^{(b)} = (u^{(b)}, V_e^{(b)})$. Obtain bootstrap realizations on the endogenous explanatory variables and the dependent variable through: $Y_e^{(b)} = Z_r\hat{\Pi}_{er} + V_e^{(b)}$ and $y^{(b)} = X^{(b)}\hat{\beta}_r + u^{(b)}$, where $X^{(b)} = (Y_e^{(b)}, Y_o, Z_1)$. Calculate the test statistic of choice Υ and store its value $\hat{\Upsilon}^{(b)}$.

2. Repeat step (1) B times resulting in the $B \times 1$ vector $\hat{\Upsilon}^B = (\hat{\Upsilon}^{(1)}, ..., \hat{\Upsilon}^{(B)})'$ of which the elements should be sorted in increasing order.

3. The null hypothesis should be rejected if for the empirical value $\hat{\Upsilon}$, calculated on the basis of y, X and Z, one finds $\hat{\Upsilon} > \hat{\Upsilon}_\alpha^{bc}$, the $(1-\alpha)(B+1)$-th value of the sorted vector.

Applying the semiparametric bootstrap is very similar as it only differs from the parametric one in step (1). Instead of assuming a distribution for the disturbances we resample by drawing rows with replacement from $\hat{U}_r$.

6.2 Simulation results for bootstrapped test statistics

Wong (1996) concludes that bootstrapping the full-set test statistics yields an improvement over using first order asymptotics, especially in the case where the (in his case external) instrument is weak. In this subsection we will discuss simulation results for the bootstrapped counterparts of the various test statistics. Again all results are obtained with $R = 10000$ and $n = 40$, additionally we choose the number of bootstrap replications to be $B = 199$. To mimic as closely as possible the way the bootstrap would be employed in practice, for each case and each test statistic we calculated the bootstrap critical value $\hat{\Upsilon}_\alpha^{bc}$ again in each separate replication.

Table 10 is the bootstrapped equivalent of Table 1. Whereas we found that the crude asymptotic version of W_o underrejects while T_o overrejects, bootstrapping these test statistics results in a substantial improvement[18] of their size properties. In fact, in this respect all three tests perform now equally well with mildly strong instruments, because the estimated actual significance level lies always inside the 99.75% confidence interval for the nominal level. Not only the sub-set tests profit from being

[18]Although the current implementation of the bootstrap already performs quite well, even better results may be obtained by rescaling the reduced form residuals by a loss of degrees of freedom correction.

bootstrapped, the one degree and two degrees of freedom full-set tests do as well. In terms of power we find that the bootstrapped versions of W_o, T_o, D_o and S_o perform almost equally well. We do find minor differences in rejection frequencies under the alternative, but often these seem still to be the results of minor differences in size. Nevertheless, on a few occasions test D_o seems to fall behind, whereas S_o seems to perform slightly better when $L > K$ (we found this too for $n = 100$). Now we establish more convincingly that exploiting correctly the exogeneity of $y^{(2)}$ in a full-set test provides more power, especially when multicollinearity is present, than not exploiting it in a sub-set test. Of course, the unfavorable substantial rejection probability of the exogeneity of the truly exogenous $y^{(3)}$, caused by wrongly treating $y^{(2)}$ as exogenous in a full-set test, cannot be healed by bootstrapping. Similar conclusions can be drawn from Table 11 which contains results for stronger instruments.

Table 10: Bootstrapped: One endogenous regressor and mildly strong instruments:

$R^2_{2;z2} = 0.20$, $R^2_{2;z23} = 0.40$, $R^2_{3;z2} = 0.20$, $R^2_{3;z23} = 0.40$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
1b	0.0	0.0	0.0	0.050	0.047	0.049	0.058	0.054	0.056	0.048	0.048	0.048	0.049	0.051	0.051	0.051	0.051	0.051	0.053	0.053
1c	0.0	0.0	0.0	0.051	0.046	0.049	0.059	0.054	0.057	0.048	0.048	0.048	0.048	0.053	0.053	0.053	0.053	0.051	0.051	0.051
2b	0.0	0.0	-0.2	0.051	0.048	0.050	0.057	0.052	0.055	0.049	0.049	0.049	0.048	0.054	0.054	0.054	0.055	0.050	0.051	0.051
2c	0.0	0.0	-0.2	0.049	0.046	0.048	0.055	0.051	0.053	0.049	0.049	0.049	0.049	0.054	0.054	0.054	0.054	0.053	0.051	0.051
3b	0.0	0.0	0.2	0.052	0.050	0.051	0.057	0.054	0.056	0.048	0.048	0.048	0.049	0.051	0.051	0.051	0.051	0.052	0.051	0.051
3c	0.0	0.0	0.2	0.052	0.048	0.050	0.056	0.053	0.055	0.049	0.049	0.049	0.049	0.051	0.051	0.051	0.051	0.052	0.052	0.052
4b	0.0	0.0	-0.5	0.051	0.051	0.051	0.055	0.053	0.055	0.049	0.049	0.049	0.049	0.052	0.052	0.052	0.052	0.051	0.051	0.051
4c	0.0	0.0	-0.5	0.051	0.051	0.050	0.055	0.053	0.054	0.047	0.047	0.047	0.047	0.052	0.052	0.052	0.052	0.051	0.050	0.050
5b	0.0	0.0	0.5	0.052	0.052	0.052	0.054	0.055	0.055	0.048	0.048	0.048	0.049	0.050	0.050	0.050	0.050	0.048	0.048	0.048
5c	0.0	0.0	0.5	0.057	0.055	0.056	0.053	0.053	0.053	0.049	0.049	0.049	0.049	0.050	0.050	0.050	0.050	0.051	0.049	0.049
6b	0.2	0.0	0.0	0.050	0.047	0.049	0.175	0.166	0.171	0.048	0.048	0.048	0.050	0.163	0.163	0.163	0.162	0.125	0.130	0.130
6c	0.2	0.0	0.0	0.050	0.047	0.049	0.172	0.166	0.171	0.051	0.051	0.051	0.053	0.163	0.163	0.163	0.163	0.124	0.128	0.128
7b	0.2	0.0	-0.2	0.051	0.048	0.051	0.170	0.155	0.168	0.084	0.084	0.084	0.085	0.194	0.194	0.194	0.195	0.155	0.152	0.152
7c	0.2	0.0	-0.2	0.049	0.048	0.048	0.170	0.158	0.167	0.079	0.079	0.079	0.081	0.198	0.198	0.198	0.198	0.156	0.154	0.154
8b	0.2	0.0	0.2	0.052	0.050	0.050	0.168	0.158	0.166	0.077	0.077	0.077	0.079	0.192	0.192	0.192	0.192	0.151	0.151	0.151
8c	0.2	0.0	0.2	0.051	0.049	0.050	0.171	0.157	0.167	0.079	0.079	0.079	0.079	0.192	0.192	0.192	0.193	0.151	0.148	0.148
9b	0.2	0.0	-0.5	0.052	0.051	0.051	0.150	0.130	0.151	0.684	0.684	0.684	0.679	0.741	0.741	0.741	0.740	0.700	0.652	0.652
9c	0.2	0.0	-0.5	0.051	0.051	0.051	0.152	0.130	0.152	0.681	0.681	0.681	0.677	0.741	0.741	0.741	0.740	0.703	0.650	0.650
10b	0.2	0.0	0.5	0.053	0.053	0.054	0.152	0.132	0.151	0.670	0.670	0.670	0.665	0.732	0.732	0.732	0.730	0.690	0.644	0.644
10c	0.2	0.0	0.5	0.054	0.055	0.054	0.150	0.131	0.150	0.675	0.675	0.675	0.670	0.733	0.733	0.733	0.731	0.693	0.649	0.649
11b	0.5	0.0	0.0	0.049	0.050	0.048	0.864	0.850	0.862	0.051	0.051	0.051	0.062	0.842	0.842	0.842	0.845	0.773	0.788	0.788
11c	0.5	0.0	0.0	0.048	0.049	0.046	0.860	0.846	0.858	0.053	0.053	0.053	0.064	0.846	0.846	0.846	0.847	0.777	0.789	0.789
12b	0.5	0.0	-0.2	0.050	0.052	0.049	0.827	0.784	0.829	0.347	0.347	0.347	0.371	0.931	0.931	0.931	0.931	0.894	0.897	0.897
12c	0.5	0.0	-0.2	0.048	0.050	0.046	0.832	0.792	0.833	0.345	0.345	0.345	0.367	0.931	0.931	0.931	0.932	0.897	0.895	0.895
13b	0.5	0.0	0.2	0.050	0.053	0.049	0.829	0.788	0.828	0.341	0.341	0.341	0.363	0.926	0.925	0.925	0.926	0.893	0.893	0.893
13c	0.5	0.0	0.2	0.048	0.052	0.047	0.822	0.781	0.821	0.345	0.345	0.345	0.370	0.928	0.928	0.928	0.929	0.896	0.896	0.896

On the other hand, we find in Table 12 that bootstrapping does not achieve satisfactory size control for most of the sub-set tests, when the instruments are weak for one regressor. And when testing the endogeneity of $y^{(2)}$ (for which the instruments are weak) there is hardly any power. The full-set tests do not show substantial size distortions and the one degree of freedom full-set test on $y^{(2)}$ and the two degrees of freedom test demonstrate power provided the regressors show multicollinearity. The results in Table 13 indicate that the sub-set test is of more use when weakness of instruments does not concern the variable under test. We can conclude that W_o and T_o have more power than D_o, since they reject less often under the null hypothesis but

Table 11: Bootstrapped: One endogenous regressor and stronger instruments:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.30$, $R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
14b	0.0	0.0	0.0	0.049	0.048	0.049	0.055	0.054	0.054	0.050	0.050	0.050	0.050	0.053	0.053	0.053	0.054	0.052	0.052	0.052
14c	0.0	0.0	0.0	0.048	0.048	0.048	0.055	0.054	0.055	0.049	0.049	0.049	0.049	0.053	0.053	0.053	0.053	0.050	0.050	0.050
15b	0.0	0.0	-0.2	0.050	0.049	0.049	0.054	0.054	0.054	0.048	0.048	0.048	0.048	0.054	0.054	0.054	0.054	0.052	0.052	0.052
15c	0.0	0.0	-0.2	0.050	0.049	0.050	0.054	0.053	0.053	0.047	0.047	0.047	0.047	0.055	0.055	0.055	0.056	0.050	0.050	0.050
16b	0.0	0.0	0.2	0.049	0.049	0.049	0.055	0.054	0.054	0.048	0.048	0.048	0.048	0.051	0.051	0.051	0.051	0.050	0.051	0.051
16c	0.0	0.0	0.2	0.053	0.052	0.052	0.055	0.055	0.055	0.048	0.048	0.048	0.048	0.053	0.053	0.053	0.052	0.051	0.049	0.049
17b	0.2	0.0	0.0	0.049	0.048	0.048	0.331	0.329	0.331	0.047	0.047	0.047	0.045	0.323	0.323	0.323	0.323	0.252	0.252	0.252
17c	0.2	0.0	0.0	0.050	0.049	0.050	0.331	0.327	0.330	0.046	0.046	0.046	0.045	0.322	0.322	0.322	0.323	0.253	0.254	0.254
18b	0.2	0.0	-0.2	0.051	0.049	0.049	0.312	0.299	0.310	0.208	0.208	0.208	0.204	0.471	0.471	0.471	0.470	0.386	0.377	0.377
18c	0.2	0.0	-0.2	0.049	0.050	0.049	0.319	0.304	0.318	0.204	0.204	0.204	0.199	0.470	0.470	0.470	0.469	0.389	0.380	0.380
19b	0.2	0.0	0.2	0.049	0.049	0.048	0.319	0.305	0.317	0.196	0.196	0.196	0.191	0.459	0.459	0.459	0.460	0.376	0.367	0.367
19c	0.2	0.0	0.2	0.052	0.052	0.051	0.314	0.299	0.312	0.204	0.204	0.204	0.199	0.458	0.458	0.458	0.457	0.378	0.369	0.369
20b	0.5	0.0	0.0	0.049	0.050	0.048	1.000	1.000	1.000	0.024	0.024	0.024	0.021	1.000	1.000	1.000	0.999	0.999	0.999	0.999
20c	0.5	0.0	0.0	0.048	0.049	0.047	0.999	0.999	0.999	0.022	0.022	0.023	0.021	0.999	0.999	0.999	0.999	0.998	0.998	0.998
21b	0.5	0.0	-0.2	0.053	0.055	0.053	0.996	0.990	0.995	0.976	0.976	0.976	0.957	1.000	1.000	1.000	1.000	1.000	1.000	1.000
21c	0.5	0.0	-0.2	0.052	0.053	0.051	0.995	0.988	0.994	0.971	0.971	0.971	0.950	1.000	1.000	1.000	1.000	1.000	1.000	1.000
22b	0.5	0.0	0.2	0.050	0.051	0.050	0.994	0.989	0.994	0.975	0.975	0.975	0.956	1.000	1.000	1.000	1.000	1.000	1.000	1.000
22c	0.5	0.0	0.2	0.053	0.056	0.053	0.994	0.989	0.994	0.977	0.977	0.977	0.957	1.000	1.000	1.000	1.000	1.000	1.000	1.000

more often under the alternative. Because we were unable yet to properly size correct the sub-set test on the strongly instrumented regressor in Tables 12 and 13, we know that we will be unable to do so too when all regressors are weakly instrumented. This is supported by the results summarized in Table 14, which also show that there is little or no power under instrument weakness.

Table 12: Bootstrapped: One endogenous regressor and weak instruments for $y^{(2)}$:

$R^2_{2;z2} = 0.01$, $R^2_{2;z23} = 0.02$, $R^2_{3;z2} = 0.30$, $R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
23b	0.0	0.0	0.0	0.030	0.037	0.029	0.032	0.053	0.048	0.049	0.049	0.049	0.048	0.050	0.050	0.050	0.050	0.048	0.049	0.049
23c	0.0	0.0	0.0	0.029	0.034	0.028	0.032	0.052	0.047	0.049	0.049	0.049	0.049	0.052	0.052	0.052	0.051	0.047	0.052	0.052
24b	0.0	0.0	0.5	0.014	0.052	0.024	0.035	0.053	0.050	0.049	0.049	0.049	0.049	0.047	0.047	0.047	0.048	0.048	0.048	0.048
24c	0.0	0.0	0.5	0.016	0.051	0.025	0.034	0.053	0.049	0.048	0.048	0.048	0.049	0.049	0.049	0.049	0.049	0.047	0.050	0.050
25b	0.2	0.0	0.0	0.034	0.037	0.033	0.033	0.056	0.050	0.051	0.051	0.051	0.051	0.053	0.053	0.053	0.054	0.049	0.051	0.051
25c	0.2	0.0	0.0	0.030	0.036	0.029	0.035	0.054	0.052	0.048	0.048	0.048	0.048	0.054	0.054	0.054	0.057	0.046	0.054	0.054
26b	0.2	0.0	0.5	0.020	0.053	0.033	0.034	0.055	0.051	0.302	0.302	0.302	0.301	0.302	0.302	0.302	0.301	0.246	0.237	0.237
26c	0.2	0.0	0.5	0.020	0.052	0.034	0.036	0.056	0.053	0.310	0.310	0.310	0.309	0.312	0.312	0.312	0.312	0.251	0.245	0.245
27b	0.5	0.0	0.0	0.034	0.044	0.033	0.045	0.078	0.068	0.048	0.048	0.048	0.048	0.063	0.063	0.063	0.066	0.044	0.067	0.067
27c	0.5	0.0	0.0	0.029	0.043	0.029	0.045	0.077	0.068	0.049	0.049	0.049	0.049	0.065	0.065	0.065	0.068	0.043	0.069	0.069
28b	0.5	0.0	0.5	0.037	0.056	0.083	0.048	0.078	0.079	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.845	1.000	1.000
28c	0.5	0.0	0.5	0.040	0.055	0.084	0.048	0.078	0.080	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.844	1.000	1.000

For DGP's in which both regressors are endogenous we again construct three tables. From subsection 5.2 we learned that under the alternative hypothesis the tests behave similar to cases in which only $y^{(2)}$ is endogenous. This is found here too as can be seen from Table 15. We find further evidence that the sub-set version of D_o performs less than W_o and T_o, whereas for some cases S_o modestly outperforms all other tests. New in comparison with Table 7 is that the two degrees of freedom full-set tests generally exhibit more power than the one degree of freedom full-set tests

Table 13: Bootstrapped: One endogenous regressor and weak instruments for $y^{(3)}$

$R^2_{2;z2} = 0.30,\ R^2_{2;z23} = 0.60,\ R^2_{3;z2} = 0.01,\ R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
29b	0.0	0.0	0.0	0.028	0.049	0.043	0.034	0.042	0.033	0.053	0.053	0.053	0.050	0.054	0.054	0.054	0.054	0.052	0.054	0.054
29c	0.0	0.0	0.0	0.030	0.047	0.045	0.034	0.038	0.034	0.052	0.052	0.052	0.053	0.054	0.054	0.054	0.054	0.053	0.052	0.052
30b	0.0	0.0	0.5	0.029	0.047	0.044	0.012	0.048	0.023	0.050	0.050	0.050	0.050	0.048	0.048	0.048	0.048	0.049	0.048	0.048
30c	0.0	0.0	0.5	0.032	0.049	0.046	0.014	0.048	0.023	0.048	0.048	0.048	0.048	0.051	0.051	0.051	0.051	0.049	0.050	0.050
31b	0.2	0.0	0.0	0.030	0.050	0.045	0.177	0.141	0.177	0.152	0.152	0.152	0.189	0.321	0.321	0.321	0.321	0.266	0.250	0.250
31c	0.2	0.0	0.0	0.029	0.048	0.044	0.178	0.136	0.180	0.157	0.157	0.157	0.190	0.320	0.320	0.320	0.320	0.257	0.249	0.249
32b	0.2	0.0	0.5	0.030	0.047	0.046	0.042	0.063	0.074	0.853	0.853	0.853	0.853	0.864	0.864	0.864	0.863	0.690	0.802	0.802
32c	0.2	0.0	0.5	0.032	0.049	0.049	0.039	0.058	0.072	0.854	0.854	0.854	0.855	0.864	0.864	0.864	0.863	0.683	0.806	0.806
33b	0.5	0.0	0.0	0.032	0.051	0.054	0.586	0.394	0.615	0.598	0.598	0.598	0.801	0.999	0.999	0.999	0.999	0.841	0.999	0.999
33c	0.5	0.0	0.0	0.027	0.050	0.053	0.589	0.397	0.615	0.598	0.598	0.598	0.799	0.999	0.999	0.999	0.999	0.843	0.999	0.999
34b	0.5	0.0	0.2	0.029	0.049	0.053	0.398	0.235	0.461	0.873	0.873	0.873	0.950	1.000	1.000	1.000	1.000	0.861	1.000	1.000
34c	0.5	0.0	0.2	0.030	0.048	0.053	0.394	0.233	0.460	0.870	0.870	0.870	0.947	1.000	1.000	1.000	1.000	0.864	1.000	1.000

Table 14: Bootstrapped: One endogenous regressor and weak instruments:

$R^2_{2;z2} = 0.01,\ R^2_{2;z23} = 0.02,\ R^2_{3;z2} = 0.01,\ R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
35b	0.0	0.0	0.0	0.028	0.035	0.028	0.029	0.038	0.029	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.052	0.050	0.051	0.051
35c	0.0	0.0	0.0	0.028	0.036	0.028	0.030	0.040	0.030	0.050	0.050	0.050	0.050	0.054	0.054	0.054	0.052	0.051	0.050	0.050
36b	0.0	0.0	0.5	0.027	0.039	0.028	0.030	0.040	0.031	0.051	0.051	0.051	0.050	0.051	0.051	0.051	0.051	0.047	0.052	0.052
36c	0.0	0.0	0.5	0.026	0.039	0.028	0.029	0.042	0.032	0.048	0.048	0.048	0.050	0.051	0.051	0.051	0.052	0.050	0.051	0.051
37b	0.2	0.0	0.0	0.027	0.037	0.028	0.030	0.038	0.030	0.053	0.053	0.053	0.053	0.051	0.051	0.051	0.052	0.051	0.053	0.053
37c	0.2	0.0	0.0	0.026	0.038	0.029	0.028	0.037	0.028	0.051	0.051	0.051	0.051	0.052	0.052	0.052	0.051	0.053	0.053	0.053
38b	0.2	0.0	0.5	0.028	0.040	0.029	0.029	0.040	0.032	0.054	0.054	0.054	0.053	0.053	0.053	0.053	0.055	0.052	0.055	0.055
38c	0.2	0.0	0.5	0.029	0.042	0.030	0.029	0.043	0.030	0.050	0.050	0.050	0.052	0.053	0.053	0.053	0.053	0.051	0.055	0.055
39b	0.5	0.0	0.0	0.026	0.044	0.029	0.031	0.048	0.034	0.062	0.062	0.062	0.063	0.061	0.061	0.061	0.065	0.054	0.066	0.066
39c	0.5	0.0	0.0	0.029	0.046	0.033	0.033	0.046	0.036	0.057	0.057	0.057	0.062	0.063	0.063	0.063	0.065	0.051	0.067	0.067
40b	0.5	0.0	0.5	0.034	0.051	0.039	0.039	0.061	0.044	0.085	0.085	0.085	0.088	0.086	0.086	0.086	0.091	0.071	0.093	0.093
40c	0.5	0.0	0.5	0.032	0.054	0.037	0.040	0.061	0.043	0.082	0.082	0.082	0.085	0.089	0.089	0.089	0.092	0.065	0.094	0.094

when the instruments are mildly strong. However, this was already found for cases with stronger instruments without bootstrapping. Increasing the instrument strength raises the rejection probabilities as before as can be seen from Table 16. That our current implementation of the bootstrap does not offer satisfactory size control for most sub-set tests when $y^{(3)}$ is weakly instrumented was already demonstrated in Table 12 and we conclude the same for the case when both regressors are endogenous as is obvious from the results in Table 17.

Table 15: Bootstrapped: Two endogenous regressors and mildly strong instruments:

$R^2_{2;z2} = 0.20,\ R^2_{2;z23} = 0.40,\ R^2_{3;z2} = 0.20,\ R^2_{3;z23} = 0.40$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W^3_2	D^3_2	T^3_2	S^3_2	W^2_3	D^2_3	T^2_3	S^2_3	W^{23}	D^{23}	T^{23}
57b	0.2	-0.2	0.0	0.164	0.156	0.161	0.176	0.168	0.173	0.161	0.161	0.161	0.164	0.167	0.167	0.167	0.171	0.218	0.230	0.230
57c	0.2	-0.2	0.0	0.174	0.164	0.170	0.174	0.168	0.172	0.164	0.164	0.164	0.168	0.170	0.170	0.170	0.173	0.223	0.234	0.234
58b	0.2	-0.2	-0.2	0.149	0.152	0.147	0.161	0.160	0.158	0.081	0.081	0.081	0.081	0.086	0.086	0.086	0.087	0.131	0.152	0.152
58c	0.2	-0.2	-0.2	0.158	0.155	0.155	0.162	0.161	0.159	0.083	0.083	0.083	0.084	0.086	0.086	0.086	0.088	0.132	0.155	0.155
59b	0.2	-0.2	0.2	0.174	0.149	0.172	0.181	0.159	0.177	0.392	0.392	0.392	0.395	0.402	0.402	0.403	0.406	0.447	0.425	0.425
59c	0.2	-0.2	0.2	0.178	0.155	0.175	0.181	0.156	0.179	0.402	0.402	0.402	0.407	0.406	0.406	0.406	0.411	0.456	0.436	0.436
60b	0.2	-0.2	-0.5	0.126	0.129	0.126	0.132	0.133	0.132	0.049	0.049	0.049	0.049	0.053	0.053	0.053	0.052	0.081	0.105	0.105
60c	0.2	-0.2	-0.5	0.131	0.131	0.130	0.134	0.134	0.132	0.050	0.050	0.050	0.049	0.053	0.053	0.053	0.052	0.079	0.106	0.106
61b	0.2	-0.2	0.5	0.171	0.128	0.171	0.168	0.129	0.170	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
61c	0.2	-0.2	0.5	0.170	0.131	0.172	0.169	0.129	0.171	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
62b	0.2	0.2	0.0	0.178	0.170	0.176	0.176	0.166	0.171	0.170	0.170	0.170	0.173	0.166	0.166	0.166	0.170	0.230	0.240	0.240
62c	0.2	0.2	0.0	0.172	0.161	0.169	0.173	0.167	0.172	0.165	0.165	0.165	0.168	0.166	0.166	0.166	0.169	0.226	0.235	0.235
63b	0.2	0.2	-0.2	0.185	0.161	0.183	0.182	0.157	0.180	0.407	0.407	0.407	0.411	0.407	0.407	0.407	0.412	0.458	0.436	0.436
63c	0.2	0.2	-0.2	0.181	0.155	0.177	0.180	0.156	0.177	0.402	0.402	0.402	0.408	0.408	0.408	0.408	0.413	0.451	0.432	0.432
64b	0.2	0.2	0.2	0.165	0.165	0.163	0.161	0.160	0.158	0.084	0.084	0.084	0.086	0.083	0.083	0.083	0.083	0.136	0.159	0.159
64c	0.2	0.2	0.2	0.159	0.160	0.158	0.161	0.160	0.159	0.084	0.084	0.084	0.086	0.081	0.081	0.081	0.083	0.132	0.153	0.153
65b	0.2	0.2	-0.5	0.172	0.134	0.174	0.169	0.126	0.172	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
65c	0.2	0.2	-0.5	0.169	0.131	0.172	0.169	0.129	0.171	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
66b	0.2	0.2	0.5	0.139	0.140	0.138	0.135	0.137	0.134	0.051	0.051	0.051	0.050	0.049	0.049	0.049	0.048	0.083	0.107	0.107
66c	0.2	0.2	0.5	0.130	0.134	0.130	0.133	0.134	0.133	0.051	0.051	0.051	0.051	0.050	0.050	0.050	0.049	0.081	0.104	0.104
67b	0.5	-0.2	0.0	0.174	0.164	0.169	0.863	0.842	0.860	0.203	0.203	0.203	0.233	0.861	0.861	0.861	0.869	0.872	0.883	0.883
67c	0.5	-0.2	0.0	0.183	0.172	0.180	0.859	0.833	0.855	0.210	0.210	0.210	0.239	0.861	0.861	0.861	0.868	0.877	0.884	0.884
68b	0.5	-0.2	-0.2	0.138	0.160	0.137	0.843	0.832	0.842	0.056	0.056	0.056	0.063	0.760	0.760	0.760	0.764	0.739	0.769	0.769
68c	0.5	-0.2	-0.2	0.145	0.167	0.144	0.847	0.835	0.845	0.055	0.055	0.055	0.064	0.755	0.754	0.754	0.759	0.738	0.771	0.771
69b	0.5	-0.2	0.2	0.203	0.155	0.197	0.815	0.735	0.816	0.887	0.887	0.887	0.900	0.998	0.998	0.998	0.998	0.998	0.999	0.999
69c	0.5	-0.2	0.2	0.204	0.157	0.200	0.807	0.727	0.808	0.887	0.887	0.887	0.902	0.997	0.997	0.997	0.997	0.998	0.998	0.998
70b	0.5	-0.2	-0.5	0.095	0.139	0.097	0.710	0.648	0.722	0.995	0.995	0.995	0.993	1.000	1.000	1.000	1.000	0.996	1.000	1.000
70c	0.5	-0.2	-0.5	0.100	0.142	0.102	0.719	0.656	0.729	0.994	0.994	0.994	0.991	1.000	1.000	1.000	1.000	0.997	1.000	1.000
71b	0.5	0.2	0.0	0.187	0.177	0.183	0.859	0.837	0.856	0.213	0.213	0.213	0.243	0.861	0.861	0.861	0.866	0.874	0.883	0.883
71c	0.5	0.2	0.0	0.182	0.169	0.178	0.863	0.841	0.860	0.204	0.204	0.204	0.234	0.866	0.866	0.866	0.872	0.874	0.881	0.881
72b	0.5	0.2	-0.2	0.208	0.162	0.205	0.813	0.728	0.813	0.889	0.889	0.889	0.902	0.997	0.997	0.997	0.997	0.997	0.998	0.998
72c	0.5	0.2	-0.2	0.206	0.159	0.203	0.818	0.736	0.817	0.884	0.884	0.884	0.897	0.996	0.996	0.996	0.996	0.998	0.999	0.999
73b	0.5	0.2	0.2	0.151	0.173	0.150	0.848	0.835	0.845	0.056	0.056	0.056	0.065	0.749	0.749	0.749	0.754	0.735	0.769	0.769
73c	0.5	0.2	0.2	0.147	0.170	0.146	0.840	0.826	0.838	0.058	0.058	0.058	0.066	0.750	0.750	0.750	0.756	0.734	0.768	0.768
74b	0.5	0.2	0.5	0.102	0.145	0.103	0.719	0.652	0.727	0.995	0.995	0.995	0.992	1.000	1.000	1.000	1.000	0.997	1.000	1.000
74c	0.5	0.2	0.5	0.097	0.143	0.099	0.709	0.647	0.717	0.995	0.995	0.995	0.992	1.000	1.000	1.000	1.000	0.996	1.000	1.000
75b	0.5	-0.5	0.0	0.859	0.788	0.854	0.861	0.790	0.856	0.944	0.944	0.944	0.958	0.948	0.948	0.948	0.959	0.999	1.000	1.000
75c	0.5	-0.5	0.0	0.861	0.796	0.857	0.856	0.782	0.852	0.945	0.945	0.945	0.959	0.948	0.948	0.948	0.960	0.998	1.000	1.000
76b	0.5	-0.5	-0.2	0.862	0.865	0.863	0.864	0.866	0.862	0.421	0.421	0.421	0.449	0.433	0.433	0.434	0.465	0.943	0.972	0.972
76c	0.5	-0.5	-0.2	0.864	0.864	0.862	0.862	0.859	0.860	0.431	0.431	0.431	0.463	0.437	0.437	0.437	0.467	0.940	0.968	0.968
77b	0.5	-0.5	-0.5	0.756	0.769	0.759	0.756	0.766	0.758	0.055	0.055	0.055	0.048	0.055	0.055	0.055	0.048	0.500	0.683	0.683
77c	0.5	-0.5	-0.5	0.761	0.776	0.764	0.761	0.773	0.766	0.054	0.054	0.054	0.049	0.051	0.051	0.051	0.045	0.503	0.686	0.686
78b	0.5	0.5	0.0	0.869	0.801	0.866	0.857	0.788	0.852	0.945	0.945	0.945	0.959	0.944	0.944	0.944	0.957	0.998	1.000	1.000
78c	0.5	0.5	0.0	0.865	0.798	0.862	0.861	0.793	0.856	0.942	0.942	0.942	0.955	0.945	0.945	0.945	0.957	0.998	1.000	1.000
79b	0.5	0.5	0.2	0.871	0.872	0.870	0.864	0.863	0.864	0.429	0.429	0.429	0.460	0.427	0.427	0.426	0.459	0.941	0.970	0.970
79c	0.5	0.5	0.2	0.869	0.868	0.867	0.866	0.864	0.865	0.430	0.430	0.430	0.459	0.430	0.430	0.430	0.461	0.943	0.970	0.970
80b	0.5	0.5	0.5	0.763	0.775	0.765	0.761	0.774	0.764	0.054	0.054	0.054	0.047	0.052	0.052	0.052	0.045	0.506	0.693	0.693
80c	0.5	0.5	0.5	0.759	0.770	0.761	0.754	0.766	0.757	0.053	0.053	0.053	0.046	0.055	0.055	0.055	0.048	0.500	0.685	0.685

Table 16: Bootstrapped: Two endogenous regressor and stronger instruments:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.30$, $R^2_{3;z23} = 0.60$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
81b	0.2	0.2	0.0	0.342	0.334	0.340	0.332	0.328	0.331	0.326	0.326	0.326	0.324	0.330	0.330	0.330	0.327	0.510	0.512	0.512
81c	0.2	0.2	0.0	0.335	0.329	0.333	0.335	0.329	0.335	0.328	0.328	0.328	0.326	0.328	0.328	0.328	0.326	0.507	0.510	0.510
82b	0.2	0.2	-0.2	0.333	0.299	0.332	0.326	0.290	0.322	0.929	0.929	0.929	0.924	0.921	0.921	0.921	0.914	0.955	0.950	0.950
82c	0.2	0.2	-0.2	0.327	0.289	0.326	0.331	0.298	0.330	0.926	0.926	0.926	0.920	0.922	0.922	0.922	0.915	0.954	0.949	0.949
83b	0.2	0.2	0.2	0.312	0.316	0.311	0.309	0.313	0.308	0.100	0.100	0.100	0.098	0.095	0.095	0.095	0.092	0.267	0.288	0.288
83c	0.2	0.2	0.2	0.304	0.308	0.302	0.304	0.308	0.303	0.096	0.096	0.096	0.092	0.096	0.096	0.096	0.093	0.269	0.289	0.289
84b	0.5	0.2	0.0	0.348	0.329	0.344	0.999	0.998	0.999	0.396	0.397	0.397	0.375	1.000	1.000	1.000	1.000	1.000	1.000	1.000
84c	0.5	0.2	0.0	0.356	0.334	0.352	0.999	0.998	0.999	0.393	0.393	0.393	0.366	1.000	1.000	1.000	1.000	1.000	1.000	1.000
85b	0.5	0.2	0.2	0.308	0.332	0.307	0.999	0.998	0.998	0.179	0.179	0.179	0.148	0.996	0.996	0.996	0.995	0.999	0.998	0.998
85c	0.5	0.2	0.2	0.298	0.324	0.297	0.999	0.998	0.999	0.184	0.184	0.184	0.149	0.996	0.996	0.996	0.995	0.999	0.999	0.999
86b	0.5	0.5	0.2	0.999	0.999	0.999	1.000	1.000	1.000	0.610	0.610	0.610	0.530	0.610	0.610	0.610	0.524	1.000	1.000	1.000
86c	0.5	0.5	0.2	1.000	1.000	1.000	1.000	0.999	1.000	0.607	0.607	0.607	0.526	0.611	0.611	0.611	0.528	1.000	1.000	1.000

Table 17: Bootstrapped: Two endogenous regressor and weak instruments for $y^{(3)}$:

$R^2_{2;z2} = 0.30$, $R^2_{2;z23} = 0.60$, $R^2_{3;z2} = 0.01$, $R^2_{3;z23} = 0.02$

Case	ρ_2	ρ_3	ρ_{23}	W^3	D^3	T^3	W^2	D^2	T^2	W_2^3	D_2^3	T_2^3	S_2^3	W_3^2	D_3^2	T_3^2	S_3^2	W^{23}	D^{23}	T^{23}
87b	0.2	0.2	0.0	0.029	0.052	0.050	0.180	0.145	0.183	0.163	0.163	0.163	0.197	0.335	0.335	0.335	0.334	0.271	0.263	0.263
87c	0.2	0.2	0.0	0.031	0.052	0.049	0.185	0.143	0.185	0.161	0.161	0.161	0.196	0.333	0.333	0.333	0.333	0.263	0.260	0.260
88b	0.2	0.2	-0.2	0.032	0.053	0.051	0.159	0.119	0.167	0.383	0.383	0.383	0.421	0.533	0.533	0.534	0.533	0.417	0.443	0.443
88c	0.2	0.2	-0.2	0.032	0.053	0.050	0.161	0.120	0.170	0.381	0.381	0.381	0.418	0.535	0.535	0.535	0.534	0.420	0.441	0.441
89b	0.2	0.2	0.2	0.030	0.051	0.048	0.089	0.098	0.096	0.162	0.162	0.162	0.179	0.249	0.249	0.249	0.249	0.206	0.195	0.195
89c	0.2	0.2	0.2	0.030	0.052	0.048	0.090	0.097	0.097	0.162	0.162	0.162	0.175	0.246	0.246	0.246	0.247	0.205	0.196	0.196
90b	0.2	0.2	-0.5	0.033	0.053	0.058	0.056	0.059	0.118	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.877	1.000	1.000
90c	0.2	0.2	-0.5	0.033	0.053	0.058	0.057	0.061	0.122	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.876	1.000	1.000
91b	0.2	0.2	0.5	0.034	0.052	0.050	0.029	0.063	0.047	0.268	0.268	0.268	0.268	0.284	0.283	0.283	0.283	0.231	0.223	0.223
91c	0.2	0.2	0.5	0.030	0.051	0.048	0.025	0.059	0.043	0.268	0.268	0.268	0.268	0.282	0.282	0.282	0.281	0.232	0.223	0.223
92b	0.5	0.2	0.0	0.032	0.053	0.057	0.583	0.391	0.614	0.614	0.614	0.614	0.814	1.000	1.000	1.000	1.000	0.841	0.999	0.999
92c	0.5	0.2	0.0	0.031	0.053	0.056	0.597	0.398	0.626	0.606	0.606	0.606	0.814	1.000	1.000	1.000	1.000	0.846	0.999	0.999
93b	0.5	0.2	-0.2	0.032	0.054	0.063	0.404	0.220	0.480	0.905	0.905	0.905	0.964	1.000	1.000	1.000	1.000	0.879	1.000	1.000
93c	0.5	0.2	-0.2	0.035	0.054	0.065	0.414	0.222	0.492	0.903	0.903	0.903	0.966	1.000	1.000	1.000	1.000	0.879	1.000	1.000
94b	0.5	0.2	0.2	0.031	0.053	0.056	0.392	0.253	0.451	0.834	0.834	0.834	0.924	0.999	0.999	0.999	0.999	0.840	0.998	0.998
94c	0.5	0.2	0.2	0.030	0.053	0.055	0.391	0.249	0.447	0.834	0.834	0.834	0.924	0.999	0.999	0.999	0.999	0.840	0.999	0.999
95b	0.5	0.5	0.0	0.047	0.078	0.085	0.604	0.387	0.641	0.683	0.683	0.683	0.867	1.000	1.000	1.000	1.000	0.851	1.000	1.000
95c	0.5	0.5	0.0	0.046	0.077	0.086	0.614	0.403	0.648	0.671	0.671	0.671	0.863	1.000	1.000	1.000	1.000	0.856	1.000	1.000
96b	0.5	0.5	0.2	0.045	0.078	0.078	0.397	0.288	0.457	0.796	0.796	0.796	0.898	0.998	0.998	0.998	0.998	0.811	0.997	0.997
96c	0.5	0.5	0.2	0.041	0.075	0.076	0.411	0.291	0.468	0.796	0.796	0.796	0.898	0.998	0.998	0.998	0.998	0.814	0.997	0.997
97b	0.5	0.5	0.5	0.049	0.082	0.078	0.104	0.136	0.210	0.998	0.998	0.998	0.998	1.000	1.000	1.000	1.000	0.838	0.998	0.998
97c	0.5	0.5	0.5	0.044	0.076	0.074	0.104	0.133	0.207	0.998	0.998	0.998	0.998	0.999	0.999	0.999	0.999	0.837	0.999	0.999

7 Empirical case study

A classic application involving more than one possibly endogenous regressor is Griliches (1976), which studies the effect of education on wage. It is often used to demonstrate instrumental variable techniques. Both education and IQ are presumably endogenous due to omitted regressors. However, testing this assumption is often overlooked. Here we shall examine the exogeneity status of both regressors jointly and individually by means of the full-set tests and the sub-set tests. The same data are used as in Hayashi

(2000, p.236). We have the wage equation and reduced form equations

$$\log W_i = \beta_1 S_i + \beta_2 IQ_i + Z_{1i}\gamma_1 + u_i \tag{7.1}$$
$$Y_i = Z_{1i}\Pi_1 + Z_{2i}\Pi_2 + V_i, \tag{7.2}$$

where W is the hourly wage rate, S is schooling in years and IQ is a test score. All explanatories of $\log W_i$ that are assumed to be predetermined or exogenous are included in Z_1; these are an intercept ($CONS$), years of experience ($EXPR$), tenure in years (TEN), a dummy for southern states (RNS) and a dummy for metropolitan areas ($SMSA$). Additionally Z_2 includes the external instruments age, age squared, mother education, KWW test score and a marital status dummy. In accordance with our previous notation both potentially endogenous regressors are included in Y.

Table 18 presents the results of four regressions. OLS treats both S and IQ as exogenous, whereas they are both assumed to be endogenous in the IV regression. In Sanderson and Windmeijer (2016) the standard first stage F-statistic is shown to be uninformative in the two-endogenous-variable model. They propose a correction to the conditional F-statistic of Angrist and Pischke (2009), which was originally introduced to deal with this issue. The corrected F-statistic[19] is 14.65 for S and 10.96 for IQ. In IV_1 regressor S is treated as predetermined and IQ as endogenous (with relevant F value of 8.77), whereas in IV_2 regressor IQ is treated as predetermined and S as endogenous (with relevant F value of 96.07). The F-values indicate that the external instruments Z_2 can be characterized as strong for S and rather weak for IQ, although in the case in which S and IQ are both treated as endogenous the results are mixed.

Table 18: Regression results for Griliches data, $n = 758$

	OLS		IV		IV_1		IV_2	
$\log W$	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
S	0.093	0.007	0.178	0.019	0.129	0.016	0.155	0.011
IQ	0.003	0.001	-0.010	0.005	-0.009	0.005	-0.002	0.001
EXPR	0.039	0.006	0.046	0.008	0.035	0.007	0.050	0.007
RNS	-0.075	0.029	-0.101	0.036	-0.110	0.034	-0.077	0.030
TEN	0.034	0.008	0.040	0.009	0.039	0.009	0.036	0.008
SMSA	0.137	0.028	0.129	0.032	0.148	0.031	0.121	0.030
CONS	30.895	0.109	40.105	0.355	40.660	0.329	30.564	0.124

Next, in Table 19, we test various hypotheses regarding the exogeneity of one or both potentially endogenous regressors. We use both the crude asymptotic tests and their refined bootstrapped versions. Joint exogeneity of schooling and IQ is rejected. Hence, at least one of these regressors is endogenous and we should use the sub-set tests to find out whether it is just one or both. However, first we examine the effect of using the full-set test on the individual regressors. In both cases the null hypothesis is rejected. From the Monte Carlo simulation results we learned

[19]The corrected F-statistics are implicitly compared to Stock-Yogo critical values.

that the full-set tests are inappropriate for correctly classifying individual regressors in the presence of other endogenous regressors. Therefore, we better employ the sub-set tests. Again we reject the null hypothesis that schooling is exogenous, but the null hypothesis that IQ is exogenous is not rejected at usual significance levels, except by the Sargan test. Bootstrapping the test statistics does not lead to different conclusions. Based on these results (and neglecting the Sargan outcome) one could greet regression IV_2 instead of IV, resulting in reduced standard errors and a less controversial result on the effect of IQ, as can be seen from Table 18. However, in some of our simulation results the Sargan test demonstrated slightly better power than its competitors. Hence, treating both S and IQ as endogenous would make sense too and yields more cautious inferences.

Table 19: DWH tests for Griliches data:

	Variables		Test Statistics					Critical Values					
Test type	Tested	Instruments	W	D	T	H	S	$\chi^2_{.05}$	$\hat{W}^{bc}_{.05}$	$\hat{D}^{bc}_{.05}$	$\hat{T}^{bc}_{.05}$	$\hat{H}^{bc}_{.05}$	$\hat{S}^{bc}_{.05}$
Full-set	S, IQ	Z_1, Z_2	46.87	59.42	65.13	40.79	66.39	5.99	6.87	7.36	7.50	6.68	7.81
Full-set	S	Z_1, Z_2, IQ	50.64	55.99	61.06	47.70	59.45	3.84	4.45	4.45	4.52	4.45	4.45
Full-set	IQ	Z_1, Z_2, S	6.28	7.24	7.38	6.23	18.58	3.84	3.32	3.56	3.61	3.31	3.62
Sub-set	S	Z_1, Z_2	41.16	45.24	46.74	38.28	47.82	3.84	5.02	5.22	5.09	4.86	5.31
Sub-set	IQ	Z_1, Z_2	2.72	3.12	2.88	2.70	6.94	3.84	3.72	4.46	4.03	3.68	4.85

8 Conclusions

In this study various tests on the orthogonality of arbitrary subsets of explanatory variables are motivated and their performance is compared in a series of Monte Carlo experiments. We find that genuine sub-set tests play an indispensable part in a comprehensive sequential strategy to classify regressors as either endogenous or exogenous. Full-set tests have a high probability to classify an exogenous regressor wrongly as endogenous if it is merely correlated with an endogenous regressor. In our derivations of the various tests we indicate flaws at various places in the established literature.

Regarding type I error performance we find that sub-set tests benefit from estimating variances under the null hypothesis (D_o), as in Lagrange multiplier tests. Estimating the variances under the alternative (W_o), as in Wald-type tests, leads to underrejection when the instruments are not very strong. However, bootstrapping results in good size control for all test statistics as long as the instruments are not weak for one of the endogenous regressors. When the various tests are compared in terms of power the bootstrapped Wald-type tests and the Sargan test behave often slightly more favorable. This falsifies earlier theoretical presumptions on the better power of the T_o type of test.

Even when the instruments are weak for the maintained endogenous regressor but strong for the regressor under inspection we find that the sub-set tests exhibit power, but there is insufficient size control, also when bootstrapped. This is in contrast to situations in which the instruments are not weak. Then, when bootstrapped, the subset and full-set tests can jointly be used fruitfully to classify individual explanatory variables and groups of them as either exogenous or endogenous. In further work we hope to report on the performance of tests of the orthogonality of sub-sets of external instruments and on joint tests of the orthogonality of included and excluded instruments.

Although our experiments involved static models of Gaussian IID variables only, with a degree of overidentification of just zero or one, we found that the nuisance parameter space which determines size distortions is quite intricate. Apart from degree of endogeneity and instrument strength, also multicollinearity and particular reduced form characteristics play a role. This may explain why so few simulation studies regarding inference on endogeneity are available yet for more realistic models.

References

Ahn, S.C., 1997. Orthogonality tests in linear models. *Oxford Bulletin of Economics and Statistics* 59, 183-186.

Angrist, J., Pischke, J.S., 2009. *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton University Press, Princeton.

Baum, C., Schaffer, M., Stillman, S., 2003. Instrumental variables and GMM: estimation and testing, *Stata Journal* 3, 1-31.

Chmelarova, V., Hill, R.C., 2010. The Hausman pretest estimator. *Economics Letters* 108, 96-99.

Davidson, R., MacKinnon, J.G., 1989. Testing for consistency using artificial regressions. *Econometric Theory* 5, 363-384.

Davidson, R., MacKinnon, J.G., 1990. Specification tests based on artificial regressions. *Journal of the American Statistical Association* 85, 220-227.

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

Doko Tchatoka, F., 2014. On bootstrap validity for specification tests with weak instruments. Forthcoming in *The Econometrics Journal*.

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

Godfrey, L.G., Hutton, J.P., 1994. Discriminating between errors-in-variables/simulaneity and misspecification in linear regression models. *Economics Letters* 44, 359-364.

Griliches, Zvi., 1976. Wages of very young men. *Journal of Political Economy* 84, S69-S85.

Guggenberger, P., 2010. The impact of a Hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory* 26, 369-382.

Hahn, J., Ham, J., Moon, H.R., 2011. The Hausman test and weak instruments. *Journal of Econometrics* 160, 289-299.

Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271.

Hausman, J.A., Taylor, W.E., 1981. A generalized specification test. *Economics*

Letters 8, 239-245.

Hayashi, Fumio, 2000. *Econometrics.* Princeton, Princeton University Press.

Holly, A., 1982. A remark on Hausman's specification test. *Econometrica* 50, 749-759.

Hwang, Hae-shin, 1980. Test of independence between a subset of stochastic regressors and disturbances. *International Economic Review* 21, 749-760.

Hwang, Hae-shin, 1985. The equivalence of Hausman and Lagrange multiplier tests of independence between disturbance and a subset of stochastic regressors. *Economics Letters* 17, 83-86.

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

Kiviet, J.F., 1985. Model selection test procedures in a single linear equation of a dynamic simultaneous system and their defects in small samples. *Journal of Econometrics* 28, 327-362.

Kiviet, J.F., 2012. *Monte Carlo Simulation for Econometricians.* Forthcoming in: Foundations and Trends in Econometrics, Now Publishers, New York.

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

Nakamura, A., Nakamura, M., 1981. On the relationships among several specification error tests presented by Durbin, Wu, and Hausman. *Econometrica* 49, 1583-88.

Nakamura, A., 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., 1985. Generalized method of moments specification testing. *Journal of Econometrics* 29, 229-256.

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

Revankar, N.S., 1978. Asymptotic relative efficiency analysis of certain tests of independence in structural systems. *International Economic Review* 19, 165-179.

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

Ruud, P.A., 1984. Tests of specification in econometrics. *Econometric Reviews* 3, 211-242 and 269-276.

Ruud, P.A., 2000. *An Introduction to Classical Econometric Theory.* Oxford University Press.

Sanderson, E., Windmeijer, F., 2016. A weak instrument F -test in linear IV models with multiple endogenous variables. *Journal of Econometrics* 190, 212-221.

Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-415.

Smith, R.J., 1983. On the classical nature of the Wu-Hausman stistics for the 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.

Spencer, D.E., Berk, K.N., 1981. A limited information specification test. *Econometrica* 49, 1079-1085 (and Erratum in *Econometrica* 50, 1087).

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Thurman, W.N., 1986. Endogeneity testing in a supply and demand framework. *The Review of Economics and Statistics* 68, 638-646.

Wong, Kafu, 1996. Bootstrapping Hausman's exogeneity test. *Economics Letters* 53, 139-143.

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 remark on a generalized specification test. *Economics Letters* 11, 365-370.