Division of Economics, EGC
School of Humanities and Social Sciences
Nanyang Technological University
14 Nanyang Drive
Singapore 637332

The Performance of Tests on Endogeneity of Subsets of Explanatory Variables Scanned by Simulation

Jan F. KIVIET and Milan PLEUS

11 December 2012

EGC Report No: 2012/08

HSS-04-88
Tel: +65 67905689
Email: D-EGC@ntu.edu.sg
http://egc.hss.ntu.edu.sg

ECONOMIC GROWTH CENTRE
Division of Economics
School of Humanities & Social Sciences

Discussion Paper

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

Jan F. Kiviet*† and Milan Pleus†

11 December 2012
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. 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. Thus, 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.

1. Introduction

In this study various test procedures are derived and examined for the classification of arbitrary subsets of explanatory variables as either endogenous or predetermined with respect to a single adequately specified structural equation. Correct classification is highly important because misclassification leads to either inefficient or inconsistent estimation. The derivations, which in essence are based on employing Hausman's principle of examining the discrepancy between two alternative estimators, formulate the various tests as joint significance tests of additional regressors in auxiliary IV regressions. Their

*Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332; jfkiviet@ntu.edu.sg.

†Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands; j.f.kiviet@uva.nl, m.pleus@uva.nl.

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. Various different and some under the null hypothesis asymptotically equivalent implementations follow. The latter vary only regarding degrees of freedom adjustments and the type of disturbance variance estimator employed. 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 of wide model classes 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.

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) and Hahn et al.(2011), 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 straight-forward 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 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

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

are finite and have full column rank, and $\hat{\beta}$ 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 further curbed 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 and not further 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 <u>e</u>ndogenous, whereas for the K_o variables Y_o their possible <u>o</u>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.

2.2. The source of an estimator discrepancy

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

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

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, its outcome could be interpreted as evidence on the validity of H^0. Because $(X'P_ZX)^{-1}$ is non-singular $\hat{\beta} - \hat{\beta}_r$ is close to zero if and only if the $K \times 1$ vector $(P_ZY_e \; P_ZY_o \; Z_1)'\hat{u}_r$ 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.

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). Upon using the well-known result that for a full column rank matrix $C = (A \; B)$ one has $P_C = P_A + P_{M_AB}$, where $M_A = I - P_A$, we find for the first vector of (2.12) -also using the first equality of (2.13)- that

$$Y_e'P_Z\hat{u}_r = Y_e'(P_{Z_r} - P_{M_ZY_o})\hat{u}_r = -Y_e'P_{M_ZY_o}\hat{u}_r.$$

[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.

Hence,

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

This K_e element vector will be close to zero when the K_o element vector $Y_o' M_Z \hat{u}_r$ is. However, 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), there are also particular circumstances possible under which $Y_e' P_Z \hat{u}_r$ will be close to zero while $Y_o' M_Z \hat{u}_r$ may not be. 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 zero.

The above shows that for all three vectors of (2.12) to be jointly close to zero, it is required that $Y_o' M_Z \hat{u}_r$ should be close to zero. This corresponds to examining to what degree the variables $M_Z Y_o$ 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 $M_Z Y_o$ the tested variables Y_o have been purged from their components spanned by the columns of Z. These are maintained to be orthogonal with respect to u, and so indeed should better be excluded from the test.

From the above it follows that $Y_o' M_Z \hat{u}_r$ being close to zero is sufficient for the full discrepancy vector (2.11) to be small. It is also necessary, since the inverse matrix in the right-hand side of (2.11) is positive definite.

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

$$y = X\beta + P_Z Y_o \zeta + u^*, \tag{2.16}$$

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

$$y = P_{Z_r} X\beta + P_{Z_r} P_Z Y_o \zeta + u^{**}. \tag{2.17}$$

Using $P_{Z_r} P_Z Y_o = P_Z Y_o$ we find for ζ the estimator

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

where $Y_o' P_Z M_{P_{Z_r} X} y = Y_o' P_Z [I - X(X' P_{Z_r} X)^{-1} X' P_{Z_r}] y = Y_o' P_Z \hat{u}_r$. Thus, by testing $\zeta = 0$ in (2.16) 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. This procedure provides the explicit solution to the exercise posed in Davidson and MacKinnon (1993, p.242).

Alternatively, consider the auxiliary regression

$$y = X\beta + M_Z Y_o \xi + v^*, \tag{2.18}$$

where $v^* = u - M_Z Y_o \xi$. Using the instruments Z_r involves here applying OLS to

$$\begin{aligned} y &= P_{Z_r} X\beta + P_{Z_r} M_Z Y_o \xi + v^{**} \\ &= P_{Z_r} X\beta + M_Z Y_o \xi + v^{**}, \end{aligned} \tag{2.19}$$

because $P_{Z_r} M_Z Y_o = P_{Z_r} Y_o - P_{Z_r} P_Z Y_o = Y_o - P_Z Y_o = M_Z Y_o$. This yields

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

where

$$\begin{aligned} Y_o' M_Z M_{P_{Z_r}X} y &= Y_o' M_Z [I - P_{Z_r} X (X' P_{Z_r} X)^{-1} X' P_{Z_r}] y \\ &= Y_o' [I - X (X' P_{Z_r} X)^{-1} X' P_{Z_r}] y - Y_o' P_Z [I - X (X' P_{Z_r} X)^{-1} X' P_{Z_r}] y \\ &= Y_o' M_Z \hat{u}_r. \end{aligned} \tag{2.21}$$

Thus, like testing $\zeta = 0$ in (2.16), testing $\xi = 0$ in auxiliary regression (2.18) examines the magnitude of $Y_o' M_Z \hat{u}_r$. This IV regression yields as the estimator for β the expression

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

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 (2.18) equals the unrestrained IV estimator $\hat{\beta}$.

From the above it follows 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 (2.18) by a Wald-type statistic, and assuming for the moment that σ^2 is known, the test statistic is

$$\sigma^{-2} y' P_{M_{P_{Z_r}X} M_Z Y_o} y = \sigma^{-2} y' (M_A - M_C) y, \tag{2.22}$$

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 (2.19). One easily finds that testing $\zeta = 0$ in (2.16) 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. This is where we have various options. 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 some kind of 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.

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

$$W_o = y'P_{M_{P_{Z_r}X}M_ZY_o}y/\hat{\sigma}^2 \text{ and } D_o = y'P_{M_{P_{Z_r}X}M_ZY_o}y/\hat{\sigma}_r^2. \tag{2.23}$$

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 $\hat{\sigma}_r^2 = y'M_Xy/n = \frac{n-K}{n}s^2$ stems from OLS. Wu (1973) suggested for this case four test statistics, indicated as $T_1, ..., T_4$, where

$$T_4 = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}D_o \text{ and } T_3 = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}W_o. \tag{2.24}$$

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

$$T_2 = \frac{T_4}{1 - \frac{K_o}{n-2K_o-L_1}T_4} = \frac{n - 2K_o - L_1}{n}\frac{1}{K_o}\frac{D_o}{1 - D_o/n}. \tag{2.25}$$

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. Because $K_e = 0$ implies $M_{P_{Z_r}X} = M_X$ we find from (2.22) that in this case

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

Hence, from the final expression we see that T_2 uses as a σ^2 estimate the OLS residual variance of auxiliary regression (2.18). Like $\hat{\sigma}^2$ and $\hat{\sigma}_r^2$, $\ddot{\sigma}^2$ is consistent under the null, because it follows from substituting (2.21) in (2.20) that $\text{plim}\,\hat{\xi} = 0$ because $\text{plim}\, n^{-1}Y_o'M_Z\hat{u}_r = 0$ under the null.

Pesaran and Smith (1990) show that under the alternative

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

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.

Next, 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 (2.18) while exploiting instruments Z_r. This amounts to the $\chi^2(K_o)$ test statistic T_o, which generalizes Wu's T_2 for cases where $K_e \geq 0$, which is given by

$$T_o = y'P_{M_{P_{Z_r}X}M_ZY_o}y/\ddot{\sigma}^2 = y'(M_A - M_C)y/\ddot{\sigma}^2, \tag{2.26}$$

with

$$\ddot{\sigma}^2 = (y - X\hat{\beta} - M_ZY_o\hat{\xi})'(y - X\hat{\beta} - M_ZY_o\hat{\xi})/n. \tag{2.27}$$

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

$$(y - X\hat{\beta} - M_Z Y_o \hat{\xi}^*)'(y - X\hat{\beta} - M_Z Y_o \hat{\xi}^*)/n \tag{2.28}$$

with

$$\hat{\xi}^* = (Y_o' M_Z Y_o)^{-1} Y_o' M_Z (y - X\hat{\beta}). \tag{2.29}$$

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 power properties. 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 $K_e = 0$).[2] Using that $\hat{\xi}$ is given by (2.29) we find from (2.27) that $\ddot{\sigma}^2 = \hat{u}' M_{M_Z Y_o} \hat{u}/n \leq \hat{\sigma}^2$, so

$$W_o \leq T_o, \tag{2.30}$$

whereas D_o can be at either side of W_o and T_o.

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. The above considerations induce that regarding tests based on the source of estimator discrepancy we will only consider the three alternative archetypical test statistics W_o, D_o and T_o, which just differ in the σ^2 estimate that they use.[3]

2.4. Testing based on the discrepancy as such

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

$$H_o = (\hat{\beta} - \hat{\beta}_r)' [\hat{\sigma}^2 (X' P_Z X)^{-1} - \hat{\sigma}_r^2 (X' P_{Z_r} X)^{-1}]^- (\hat{\beta} - \hat{\beta}_r), \tag{2.31}$$

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, (2.31) 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.

[2] 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 (2.23). 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.

[3] 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. All three 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.

Let β contain components as follows from the decompositions

$$X\beta = Y_e\beta_e + Y_o\beta_o + Z_1\beta_1 = Y\beta_{eo} + Z_1\beta_1, \tag{2.32}$$

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

$$\begin{aligned} \hat{\beta} - \hat{\beta}_r &= (X'P_ZX)^{-1} \begin{pmatrix} Y'P_Z\hat{u}_r \\ 0 \end{pmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix} Y'P_Z\hat{u}_r, \\ \hat{\beta}_{eo} - \hat{\beta}_{eo,r} &= A_{11}Y'P_Z\hat{u}_r. \end{aligned} \tag{2.33}$$

Hence, the discrepancy vector of the two coefficient estimates of just the regressors in Y, but also 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 (2.31) can also be obtained from

$$H_o = (\hat{\beta}_{eo} - \hat{\beta}_{eo,r})'[\hat{\sigma}^2(Y'P_{M_{Z_1}Z_2}Y)^{-1} - \hat{\sigma}_r^2(Y'P_{M_{Z_1}(Z_2\ Y_o)}Y)^{-1}]^-(\hat{\beta}_{eo} - \hat{\beta}_{eo,r}). \tag{2.34}$$

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

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

Calculating (2.34) instead of (2.31) 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 an inverse 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[4].

Since (2.20) and (2.21) give $(Y_o'M_ZM_{P_{Z_r}X}M_ZY_o)\hat{\xi} = Y_o'M_Z\hat{u}_r$ we can establish the linear transformation

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

which indicates 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[5].

[4]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 (2.31) and (2.34).

[5]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

In line with auxiliary regression (2.18), we can examine independence between u and Y_o or its reduced form disturbances directly, instead of proxying the latter by the residuals $M_Z Y_o$. Consider regression (2.1) augmented by the actual reduced form disturbances

$$y = X\beta + (Y_o - Z\Pi_o)\phi + w^*, \tag{2.35}$$

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 (2.35) can be written as

$$\begin{aligned} y &= Y_e\beta_e + Y_o(\beta_o + \phi) + Z_1(\beta_1 - \Pi_{o1}\phi) - Z_2\Pi_{o2}\phi + w^* \\ &= X\beta^* + Z_2\phi^* + w^* \end{aligned} \tag{2.36}$$

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 (2.35), through estimating (2.36) consistently is not an option, unless $K_e = 0$. For $K_e > 0$, which is the case of our primary interest here, (2.36) 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.[6] 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 (2.36) and the IV estimator of β_o.[7]

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 (2.36) 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. This yields the familiar Sargan-Hansen test for testing the overidentification restrictions of model (2.1). It is well-known that this test has power for alternatives in which some of the variables in Z_2 are actually omitted regressors (or are correlated with u). In practical situations this type of test, and also Hausman type tests for the orthogonality of particular instruments not included in the specification[8], are very useful. However, we do not consider them here, because right from the beginning we have chosen a context in which all instruments Z are assumed to be uncorrelated with u.

2.6. Testing by an incremental Sargan test

The test of overidentifying restrictions initiated by Sargan (1958) does not enable to infer directly on the orthogonality of individual instrumental variables, but an incremental

[6] 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.

[7] 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.

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

Sargan test builds on the maintained hypothesis $E(Z'u) = 0$ and can test the orthogonality of additional potential instrumental variables. Choosing for these the regressors Y_o yields a test statistic for the hypotheses (2.9) which is given by

$$S_o = \frac{\hat{u}_r' P_{Z_r} \hat{u}_r}{\hat{\sigma}_r^2} - \frac{\hat{u}' P_Z \hat{u}}{\hat{\sigma}^2}. \tag{2.37}$$

When using for both separate Sargan statistics the same σ^2 estimate 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 (2.22) is given by $y'(P_{(P_{Z_r}X \; M_Z Y_o)} - P_{P_{Z_r}X})y$. We shall prove equivalence[9] by using the general results for the full column rank matrices $C = (A \; B)$ and $C^* = (A \; B^*)$, where $B^* = B + AD$ and D is an arbitrary matrix of appropriate dimensions, that not only (i) $P_C = P_{(A \; M_A B)} = P_A + P_{M_A B}$, but also (ii) $P_{C^*} = P_A + P_{M_A B^*} = P_A + P_{M_A B} = P_C$. Using (i) we have $P_{Z_r}X = P_Z X + P_{M_Z Y_0} X$, which gives

$$P_{(P_{Z_r}X \; M_Z Y_o)} = P_{(P_Z X + P_{M_Z Y_0} X \; M_Z Y_o)} = P_{(P_Z X + M_Z Y_0 X^* \; M_Z Y_o)},$$

where $X^* = (Y_0' M_Z Y_0)^{-1} Y_0' M_Z X$. Next by (ii) we obtain

$$P_{(P_Z X + M_Z Y_0 X^* \; M_Z Y_o)} = P_{(P_Z X \; M_Z Y_o)},$$

which completes the proof.

2.7. Preliminary selection

The foregoing subsections demonstrate that all available statistics W_o, D_o, T_o, H_0 and S_o for testing the orthogonality of a subset of the potentially endogenous regressors basically just differ regarding the estimation of σ^2. Both S_o and H_o show a hybrid nature in this respect, because their most natural implementations require two different σ^2 estimates, which may lead to negative test outcomes. In addition to that H_o has the drawback that it involves a generalized inverse. Similar differences and correspondences carry over to more general models, which would require GMM estimation, 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. Therefore, in the remainder of this study, we will examine the performance in finite samples of the three archetypical tests W_o, D_o and T_o. Because S_o and H_o are in fact hybridly weighted versions of these we do not expect them to have unique qualities that deserve substantial attention.

[9] 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.

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 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 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 (2.26), 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.[10] 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[11]. Their Monte Carlo design is very restricted, 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[12] 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.

[10] 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.

[11] 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 but always perform tests based on estimator contrasts by running the relevant auxiliary regression.

[12] 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.

4.1. The simulated data generating process

Model (2.1) is specialized in our simulations to

$$y = \beta_1 \iota + \beta_2 y^{(2)} + \beta_3 y^{(3)} + u, \tag{4.1}$$
$$y^{(2)} = \pi_{21}\iota + \pi_{22} z^{(2)} + \pi_{23} z^{(3)} + v^{(2)}, \tag{4.2}$$
$$y^{(3)} = \pi_{31}\iota + \pi_{32} z^{(2)} + \pi_{33} z^{(3)} + v^{(3)}, \tag{4.3}$$

where ι is an $n \times 1$ vector consisting of ones. So, $K = 3$, $L_1 = 1$ and $L_2 = 2$, with $K_o + K_e = 2$, $Y = (y^{(2)}\ y^{(3)})$, $Z_1 = \iota$ and $Z = (\iota\ z^{(2)}\ z^{(3)})$. Since $K = L$, at this stage we only investigate the case in which under the unrestrained alternative hypothesis the single simultaneous equation (4.1) is just identified according to the order condition. Because the statistics to be analyzed will be invariant regarding the values of the intercepts, these are all set equal to zero, thus $\beta_1 = \pi_{21} = \pi_{31} = 0$. Fulfillment of the rank condition for identification then implies that the inequality

$$\pi_{22}\pi_{33} \neq \pi_{23}\pi_{32} \tag{4.4}$$

has to be satisfied.

The vectors $z^{(2)}$ and $z^{(3)}$ will be generated as mutually independent $IID(0, 1)$ series. They have been drawn only once and then were kept fixed over all replications. In fact we drew two arbitrary series and next 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 have components

$$v^{(2)} = \eta^{(2)} + \gamma_2 u \text{ and } v^{(3)} = \eta^{(3)} + \kappa\eta^{(2)} + \gamma_3 u, \tag{4.5}$$

where the series u_i, $\eta_i^{(2)}$ and $\eta_i^{(3)}$ will be generated as mutually independent zero mean IID series (for $i = 1, ..., n$). Without loss of generality, we may choose $\sigma_u^2 = 1$. 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 will realize $\sigma^2_{y^{(2)}} = \sigma^2_{y^{(3)}} = 1$ 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

$$\begin{aligned} \sigma^2_{v^{(2)}} &= \sigma^2_{\eta^{(2)}} + \gamma_2^2, & \sigma^2_{y^{(2)}} &= \pi_{22}^2 + \pi_{23}^2 + \sigma^2_{v^{(2)}} = 1, \\ \sigma^2_{v^{(3)}} &= \sigma^2_{\eta^{(3)}} + \kappa^2\sigma^2_{\eta^{(2)}} + \gamma_3^2, & \sigma^2_{y^{(3)}} &= \pi_{32}^2 + \pi_{33}^2 + \sigma^2_{v^{(3)}} = 1. \end{aligned} \tag{4.6}$$

This requires

$$\sigma^2_{\eta^{(2)}} = 1 - \pi_{22}^2 - \pi_{23}^2 - \gamma_2^2 > 0 \text{ and } \sigma^2_{\eta^{(3)}} = 1 - \pi_{32}^2 - \pi_{33}^2 - \kappa^2\sigma^2_{\eta^{(2)}} - \gamma_3^2 > 0. \tag{4.7}$$

In addition to (4.4), (4.7) implies two further inequality restrictions on the nine parameters of the data generating process, which are

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

However, more restrictions should be respected as we will see, when we consider further consequences of a choice of particular values for these DGP parameters.

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

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

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

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

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

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

This yields

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

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 $\pi^2_{22} \geq \pi^2_{23}$. 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

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

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 (4.4) and (4.7). Moreover, we will show that not all of these 2^4 solutions necessarily lead to unique results on the distribution of the test statistics W_o, D_o and T_o.

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

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

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 the three sub-set test statistics are all 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 $\ddot{\sigma}^2$ is too, because $y-X\hat{\beta}-M_ZY_o\hat{\xi}=\hat{u}-M_ZY_o\hat{\xi}$ is, as follows from (2.20) and (2.21). Moreover, because $\hat{\xi}$ is invariant with respect to β also is the numerator of the three test statistics.[13] Therefore, $R^2_{1;z2}$ and $R^2_{1;z23}$ do not really establish nuisance parameters, reducing the dimensionality of the nuisance parameter space to 7. Without loss of generality we may always 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 test statistics W_o, D_o and T_o, because it does not alter the space spanned by the matrices A and C of (2.22) 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.

For the design parameters we shall choose various interesting combinations from

[13] 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.

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

in as far as they satisfy the restrictions (4.18) through (4.21), provided they obey also the admissibility restrictions given by (4.4), (4.7) 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 $\eta^{(3)}$. The three test statistics W_o, D_o and T_o will be calculated for both $y^{(2)}$ (then denoted as W^2, D^2, T^2) and for $y^{(3)}$ (denoted W^3, D^3, T^3) assuming the other regressor to be endogenous. These genuine sub-set tests will be compared with tests on the endogeneity of the full set. The latter are denoted W^{23}, D^{23}, T^{23} (these are tests involving 2 degrees of freedom), W^2_3, D^2_3, T^2_3 (when $y^{(3)}$ is treated as exogenous) and W^3_2, D^3_2, T^3_2 (when $y^{(2)}$ is treated as exogenous). The behavior under both the null and the alternative hypothesis will be investigated. These full-set tests are included to better appreciate the special nature of the more subtle sub-set tests under investigation here.

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

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

Here $\mathbb{I}(.)$ 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

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

We choose $R = 10000$, examine all endogeneity tests at the nominal significance level of 5%, and take the sample size equal to $n = 40$ (mostly). 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.

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 three 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 shows acceptable size control, whereas W_o tends to underreject, whilst T_o overrejects. 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, D^3 and T^3. 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, D^2 and T^2 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 exploits genuinely exogenous regressors, because if it does not it may falsely diagnose endogeneity of an exogenous regressor (but by a reasonably low probability when the regressors are uncorrelated). However, the next tests reported, which exploit the genuine exogeneity of $y^{(3)}$, demonstrate that in this case they 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 (2.16), 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. One would expect lower rejection probability here than for the full-set test which correctly exploits orthogonality of $y^{(3)}$, but 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 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.

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 (4.7). 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.

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 sub-set tests examine different situations (even when $\rho_2 = \rho_3 = 0$) and so do the two one degree of freedom full-set tests. 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. However, the probability that sub-set test D^2 will detect the endogeneity is small, whereas that of D_3^2 is better only under multicollinearity. D_2^3 will again provide confusing evidence, unless the regressors are orthogonal.

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 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 instruments) and W_2^3 (provided the multicollinearity is substantial). The full-set T_o tests are again oversized. All D_o implementations show mild size distortions. Sub-set test D^2 has power especially when the regressors show little multicollinearity, but after size correction it seems likely that W^2 or especially T^2 would do much better. Also the tests W_3^2, D_3^2 and T_3^2 show power for detecting endogeneity of $y^{(2)}$ when the instruments are weak for exogenous regressor $y^{(3)}$, and their power increases with multicollinearity.

Finally we construct DGP's in which the instruments are weak for both regressors. 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 size correction, although it will yield very modest power.

In addition to cases in which the two instruments have similar strength for $y^{(2)}$ and $y^{(3)}$, we present a couple of cases in which this differs. Note that the inequality (4.4) 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 (4.7). The results for the sub-set tests differ greatly between the four subcases. Subcases "a" and "d" show lower rejection probabilities for W_o and T_o, whereas D_o seems unaffected 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 similar 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 treated as 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 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 multicollinearity 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 multicollinearity 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 two 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.

Because Table 3 and 4 are very similar and now both regressors are endogenous we only need to consider the equivalent table of the latter. 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, though we do not present a table on these as it is not very informative.

These simulations demonstrate that the sub-set tests are indispensable when there is more than one regressor in a model that might be endogenous. Using only full-set tests will not enable to classify the individual variables as either endogenous or exogenous. However, all tests examined here 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 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[14] 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 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 and D_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. 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.

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. Only D^2 shows reasonable type I error probabilities, but when testing the endogeneity of $y^{(2)}$, the regressor 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 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. Again the results are slightly better for W_o and T_o but there is almost no power.

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

[14] Although the current implementation of the bootstrap already performs quite well, even better results may be obtained by rescaling the reduced form residuals.

than W_o and T_o. 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 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.

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

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

where W is the hourly wage rate, S is schooling in years and IQ is a test score. All regressors 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 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 schooling and IQ as exogenous, whereas they are assumed to be endogenous in the IV regression. In IV_1 only IQ is treated as endogenous whereas in IV_2 only schooling is treated as endogenous.

Next, in Table 19, we test various hypotheses regarding the exogeneity of one or both potentially endogenous regressors. 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 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 should not be rejected. Bootstrapping these two test statistics does not lead to different conclusions. Based on these results one should 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.

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.

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 behave often more favorable. This falsifies earlier theoretical presumptions on the better power of the T_o type of test. The outcome is such that we do not expect that a better performance could have been obtained by the computationally more involved implementations that result from strictly employing the Hausman or the Hansen-Sargan principles.

Even when the instruments are weak for the maintained endogenous regressor but strong for the regressor under inspection we find that the auxiliary regression 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 sub-set and full-set tests can jointly be used fruitfully to classify individual explanatory variables and groups of them as either exogenous or endogenous.

It must be noted though that the conclusions obtained from the experiments in this study are limited, as they only deal with static linear models with Gaussian disturbances, which are just identified by genuinely exogenous external instruments. Apart from relaxing some of these limitations in future work, we plan to look further into effects due to weakness of instruments. Furthermore, tests on the orthogonality of sub-sets of external instruments and joint tests on the orthogonality of included and excluded instruments deserve further examination.

References

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

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.

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

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

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.

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.

Table 1: One endogenous regressor and mildly strong instruments:
$R^2_{2;z2} = .2,\ R^2_{2;z23} = .4,\ R^2_{3;z2} = .2,\ R^2_{3;z23} = .4$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
1b	0	0	0	.033	.054	.064	.030	.050	.061	.040	.050	.070	.040	.046	.068	.021	.044	.088
1c	0	0	0	.036	.056	.069	.030	.053	.063	.042	.050	.073	.037	.047	.069	.023	.046	.086
2b	0	0	-.2	.032	.055	.064	.029	.055	.064	.044	.050	.069	.043	.047	.067	.023	.044	.090
2c	0	0	-.2	.034	.057	.067	.031	.053	.062	.044	.050	.074	.040	.048	.068	.024	.047	.086
3b	0	0	.2	.034	.056	.065	.028	.051	.061	.045	.049	.069	.041	.046	.068	.023	.044	.087
3c	0	0	.2	.035	.057	.069	.029	.053	.063	.047	.052	.073	.041	.046	.068	.023	.044	.086
4b	0	0	-.5	.024	.058	.069	.024	.057	.068	.057	.052	.073	.058	.052	.073	.027	.046	.090
4c	0	0	-.5	.024	.056	.067	.023	.056	.066	.059	.052	.076	.054	.050	.072	.029	.047	.090
5b	0	0	.5	.026	.060	.070	.023	.056	.067	.054	.048	.071	.053	.048	.071	.027	.045	.083
5c	0	0	.5	.025	.057	.066	.023	.056	.067	.057	.052	.075	.055	.049	.070	.028	.047	.086
6b	.2	0	0	.033	.056	.064	.122	.177	.199	.039	.047	.074	.134	.161	.208	.067	.128	.200
6c	.2	0	0	.036	.058	.067	.125	.180	.203	.043	.051	.074	.135	.157	.205	.067	.128	.202
7b	.2	0	-.2	.032	.057	.064	.120	.174	.198	.068	.076	.108	.172	.186	.239	.083	.147	.228
7c	.2	0	-.2	.032	.059	.067	.122	.174	.198	.075	.085	.117	.177	.192	.243	.088	.148	.233
8b	.2	0	.2	.034	.058	.064	.117	.168	.197	.076	.085	.116	.177	.195	.246	.090	.150	.231
8c	.2	0	.2	.034	.059	.067	.122	.173	.198	.072	.082	.112	.172	.190	.243	.090	.150	.230
9b	.2	0	-.5	.025	.059	.067	.103	.155	.192	.701	.682	.743	.759	.744	.798	.609	.644	.747
9c	.2	0	-.5	.026	.058	.066	.092	.143	.183	.705	.687	.744	.763	.745	.800	.609	.645	.748
10b	.2	0	.5	.027	.060	.071	.093	.144	.182	.709	.691	.754	.767	.751	.804	.614	.651	.755
10c	.2	0	.5	.025	.058	.066	.102	.154	.192	.707	.690	.748	.763	.746	.800	.612	.646	.753
11b	.5	0	0	.028	.063	.059	.816	.865	.885	.041	.051	.075	.825	.848	.886	.634	.782	.861
11c	.5	0	0	.031	.065	.064	.810	.865	.888	.042	.052	.075	.822	.846	.884	.638	.783	.858
12b	.5	0	-.2	.028	.064	.060	.768	.818	.862	.322	.344	.410	.929	.933	.952	.814	.895	.939
12c	.5	0	-.2	.030	.065	.064	.763	.809	.850	.333	.355	.421	.930	.934	.954	.820	.898	.941
13b	.5	0	.2	.031	.065	.062	.762	.808	.852	.335	.358	.425	.934	.938	.958	.818	.898	.944
13c	.5	0	.2	.030	.066	.063	.771	.816	.859	.321	.343	.406	.932	.936	.956	.815	.900	.943

Table 2: One endogenous regressor and stronger instruments:
$R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
14b	0	0	0	.048	.058	.068	.042	.052	.065	.052	.051	.070	.046	.046	.068	.039	.044	.089
14c	0	0	0	.052	.061	.072	.045	.056	.068	.055	.053	.076	.048	.047	.070	.042	.047	.086
15b	0	0	-.2	.045	.058	.068	.045	.057	.070	.054	.050	.071	.052	.048	.070	.042	.045	.092
15c	0	0	-.2	.048	.059	.070	.045	.055	.068	.056	.052	.076	.051	.046	.070	.042	.047	.087
16b	0	0	.2	.048	.060	.070	.042	.054	.065	.053	.049	.070	.050	.047	.069	.040	.044	.087
16c	0	0	.2	.046	.058	.072	.044	.056	.068	.057	.053	.076	.051	.045	.066	.042	.046	.088
17b	.2	0	0	.049	.058	.068	.328	.358	.392	.047	.046	.066	.329	.325	.387	.224	.241	.348
17c	.2	0	0	.051	.061	.072	.329	.356	.391	.049	.047	.067	.328	.324	.388	.229	.247	.353
18b	.2	0	-.2	.045	.058	.069	.306	.331	.372	.213	.202	.258	.482	.468	.535	.363	.372	.491
18c	.2	0	-.2	.046	.059	.071	.304	.329	.370	.220	.210	.260	.485	.471	.536	.361	.367	.488
19b	.2	0	.2	.047	.061	.070	.303	.326	.370	.225	.212	.270	.478	.464	.538	.361	.368	.488
19c	.2	0	.2	.046	.059	.071	.307	.334	.375	.219	.208	.259	.480	.464	.534	.359	.365	.483
20b	.5	0	0	.045	.063	.066	1	1	1	.023	.023	.037	1	1	1	.999	.999	1
20c	.5	0	0	.048	.066	.069	1	1	1	.025	.024	.039	1	1	1	.999	.999	1
21b	.5	0	-.2	.043	.065	.067	.994	.992	.996	.978	.975	.987	1	1	1	1	1	1
21c	.5	0	-.2	.041	.061	.065	.994	.993	.995	.981	.978	.988	1	1	1	1	1	1
22b	.5	0	.2	.043	.061	.065	.994	.993	.996	.980	.977	.987	1	1	1	1	1	1
22c	.5	0	.2	.040	.061	.064	.994	.992	.996	.980	.976	.987	1	1	1	1	1	1

Table 3: One endogenous regressor and weak instruments for $y^{(2)}$:
$R^2_{2;z2} = .01,\ R^2_{2;z23} = .02,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
23a	0	0	0	.013	.031	.021	.001	.054	.051	.051	.050	.072	.001	.048	.070	.005	.044	.086
23b	0	0	0	.014	.032	.023	0	.054	.050	.055	.055	.076	.002	.047	.070	.005	.046	.089
24a	0	0	.5	.001	.052	.012	.001	.054	.056	.056	.050	.071	.047	.049	.071	.008	.045	.084
24b	0	0	.5	.001	.054	.016	.001	.055	.057	.060	.053	.076	.051	.052	.075	.007	.047	.087
25a	.2	0	0	.013	.032	.021	.001	.057	.053	.050	.048	.071	.001	.047	.072	.005	.047	.089
25b	.2	0	0	.011	.033	.021	0	.058	.054	.052	.052	.075	.002	.049	.070	.006	.047	.092
26a	.2	0	.5	.003	.053	.026	.001	.059	.061	.345	.321	.389	.317	.324	.390	.083	.248	.353
26b	.2	0	.5	.004	.056	.026	0	.060	.062	.337	.317	.381	.309	.316	.383	.083	.238	.342
27a	.5	0	0	.012	.039	.019	.002	.089	.080	.049	.047	.069	.002	.062	.090	.005	.061	.113
27b	.5	0	0	.012	.040	.020	.002	.087	.081	.050	.048	.069	.003	.059	.086	.005	.063	.114
28a	.5	0	.5	.023	.060	.126	.002	.091	.103	1	1	1	1	1	1	.694	1	1
28b	.5	0	.5	.023	.062	.122	.002	.091	.109	1	1	1	1	1	1	.696	1	1

Table 4: One endogenous regressor and weak instruments for $y^{(3)}$:
$R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .01,\ R^2_{3;z23} = .02$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
29b	0	0	0	.001	.058	.054	.012	.032	.020	.002	.045	.067	.048	.046	.069	.004	.044	.088
29c	0	0	0	.001	.058	.055	.011	.031	.021	.001	.049	.074	.049	.046	.070	.004	.044	.088
30b	0	0	.5	.001	.057	.060	.002	.056	.012	.047	.047	.070	.053	.047	.069	.007	.044	.087
30c	0	0	.5	.001	.060	.062	.001	.058	.013	.046	.048	.069	.054	.047	.068	.006	.045	.088
31b	.2	0	0	.001	.059	.056	.100	.137	.141	.006	.156	.201	.328	.324	.386	.068	.242	.347
31c	.2	0	0	.001	.059	.057	.101	.141	.145	.005	.156	.197	.329	.325	.384	.070	.244	.349
32b	.2	0	.5	.001	.058	.069	.017	.072	.086	.867	.869	.904	.891	.880	.912	.428	.810	.881
32c	.2	0	.5	.001	.061	.069	.016	.072	.094	.866	.868	.901	.892	.880	.910	.433	.809	.878
33b	.5	0	0	.001	.063	.074	.572	.460	.643	.068	.600	.630	1	1	1	.667	.999	1
33c	.5	0	0	.001	.064	.075	.570	.460	.637	.068	.593	.626	1	1	1	.658	.999	1
34b	.5	0	.2	.001	.063	.074	.414	.298	.536	.429	.870	.883	1	1	1	.699	1	1
34c	.5	0	.2	.001	.064	.081	.420	.311	.543	.430	.868	.884	1	1	1	.700	1	1

Table 5: One endogenous regressor and weak instruments:
$R^2_{2;z2} = R^2_{3;z2} = .01,\ R^2_{2;z23} = R^2_{3;z23} = .02$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
35b	0	0	0	0	.033	.016	0	.032	.017	.002	.046	.068	.002	.047	.068	0	.043	.084
35c	0	0	0	0	.034	.016	0	.031	.016	.002	.047	.070	.002	.047	.067	0	.044	.083
36b	0	0	.5	0	.035	.020	.001	.037	.018	.003	.046	.069	.002	.046	.065	0	.044	.086
36c	0	0	.5	0	.033	.017	0	.034	.017	.002	.048	.068	.002	.048	.070	0	.044	.083
37b	.2	0	0	0	.033	.017	0	.033	.016	.002	.050	.071	.001	.047	.068	0	.046	.089
37c	.2	0	0	0	.033	.016	0	.032	.017	.002	.050	.071	.001	.049	.070	0	.047	.090
38b	.2	0	.5	0	.037	.019	0	.036	.018	.003	.049	.074	.002	.050	.075	0	.050	.095
38c	.2	0	.5	0	.038	.018	0	.036	.017	.002	.052	.076	.002	.054	.079	0	.049	.092
39b	.5	0	0	0	.037	.018	0	.042	.022	.002	.058	.086	.003	.063	.088	0	.060	.112
39c	.5	0	0	0	.038	.017	.001	.041	.021	.002	.057	.084	.003	.061	.089	0	.059	.112
40b	.5	0	.5	0	.051	.026	0	.054	.032	.005	.086	.118	.007	.095	.125	0	.093	.157
40c	.5	0	.5	0	.050	.024	.001	.057	.031	.005	.082	.115	.005	.091	.125	0	.092	.152

Table 6: One endogenous regressor and asymmetric instrument strength: $R^2_{2;z2} = .3,\ R^2_{2;z23} = .5,\ R^2_{3;z2} = .1,\ R^2_{3;z23} = .4$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
1a	0	0	0	.001	.059	.036	.001	.057	.028	.057	.051	.072	.057	.050	.073	.008	.048	.090
1b	0	0	0	.034	.056	.066	.037	.053	.063	.041	.050	.069	.045	.048	.070	.026	.045	.089
1c	0	0	0	.033	.059	.071	.036	.052	.064	.043	.052	.073	.044	.047	.068	.027	.043	.086
1d	0	0	0	.002	.060	.036	.001	.059	.028	.054	.048	.069	.054	.049	.070	.008	.047	.089
2b	0	0	-.2	.030	.059	.069	.031	.058	.066	.049	.050	.073	.050	.048	.069	.027	.045	.093
2c	0	0	-.2	.037	.057	.070	.038	.053	.064	.043	.051	.073	.045	.047	.071	.028	.045	.086
2d	0	0	-.2	.002	.058	.024	.002	.057	.018	.043	.047	.069	.047	.047	.069	.006	.044	.089
3a	0	0	.2	.002	.058	.025	.003	.056	.018	.044	.049	.073	.048	.048	.071	.008	.045	.088
3b	0	0	.2	.037	.056	.068	.036	.049	.061	.041	.051	.071	.044	.046	.073	.027	.044	.088
3c	0	0	.2	.033	.059	.068	.030	.055	.063	.049	.050	.073	.047	.045	.067	.025	.045	.088
4c	0	0	-.5	.032	.055	.069	.033	.054	.066	.056	.052	.075	.054	.048	.071	.033	.046	.089
4d	0	0	-.5	.004	.054	.023	.006	.050	.022	.020	.047	.070	.035	.048	.071	.006	.044	.088
5a	0	0	.5	.005	.054	.026	.007	.047	.022	.022	.050	.071	.032	.046	.068	.007	.044	.089
5b	0	0	.5	.034	.060	.071	.030	.054	.066	.054	.049	.069	.054	.048	.069	.031	.043	.085
6a	.2	0	0	.002	.059	.044	.005	.073	.073	.703	.684	.742	.714	.693	.753	.304	.590	.701
6b	.2	0	0	.033	.057	.066	.190	.233	.264	.050	.059	.087	.224	.231	.289	.123	.177	.265
6c	.2	0	0	.034	.059	.069	.192	.234	.265	.053	.062	.087	.224	.230	.288	.128	.179	.267
6d	.2	0	0	.003	.059	.046	.005	.072	.074	.710	.692	.751	.721	.700	.758	.310	.596	.706
7b	.2	0	-.2	.030	.060	.068	.161	.206	.245	.186	.189	.246	.349	.343	.410	.207	.258	.370
7c	.2	0	-.2	.035	.059	.069	.203	.240	.272	.053	.062	.090	.231	.239	.301	.134	.185	.275
7d	.2	0	-.2	.004	.058	.033	.010	.079	.056	.249	.265	.322	.285	.287	.348	.080	.223	.325
8a	.2	0	.2	.003	.059	.032	.009	.080	.057	.246	.264	.322	.284	.286	.349	.082	.220	.317
8b	.2	0	.2	.036	.058	.067	.202	.239	.272	.055	.066	.093	.232	.239	.298	.134	.182	.275
8c	.2	0	.2	.031	.060	.068	.162	.206	.245	.188	.192	.246	.343	.337	.406	.217	.266	.369
9c	.2	0	-.5	.031	.056	.068	.166	.200	.250	.615	.599	.665	.724	.708	.763	.586	.604	.710
9d	.2	0	-.5	.005	.056	.032	.030	.104	.076	.056	.115	.151	.134	.172	.218	.037	.139	.214
10a	.2	0	.5	.008	.056	.031	.035	.107	.077	.055	.114	.147	.130	.167	.215	.038	.135	.211
10b	.2	0	.5	.033	.061	.072	.164	.198	.248	.622	.608	.670	.729	.715	.771	.589	.607	.720
11b	.5	0	0	.032	.062	.064	.954	.958	.972	.145	.165	.208	.979	.979	.987	.937	.960	.981
11c	.5	0	0	.032	.064	.066	.955	.959	.973	.147	.165	.213	.980	.979	.987	.942	.963	.983
12b	.5	0	-.2	.030	.065	.064	.853	.848	.911	.930	.930	.951	1	1	1	.996	1	1
12c	.5	0	-.2	.032	.064	.066	.960	.963	.973	.156	.176	.219	.986	.986	.990	.960	.971	.987
12d	.5	0	-.2	.014	.063	.089	.100	.201	.330	.989	.991	.995	.997	.997	.998	.701	.995	.998
13a	.5	0	.2	.013	.064	.089	.102	.197	.328	.991	.992	.995	.997	.996	.998	.699	.994	.998
13b	.5	0	.2	.032	.063	.063	.961	.963	.975	.159	.179	.228	.986	.986	.992	.963	.973	.987
13c	.5	0	.2	.030	.063	.063	.850	.843	.912	.930	.930	.951	1	1	1	.996	1	1
14d	.5	0	-.5	.025	.066	.074	.302	.432	.498	.438	.619	.674	.867	.909	.934	.495	.872	.926
15a	.5	0	.5	.022	.067	.074	.300	.427	.497	.443	.627	.681	.871	.910	.934	.493	.873	.926

Table 7: Two endogenous regressors and mildly strong instruments:
$R^2_{2;z2} = .2,\ R^2_{2;z23} = .4,\ R^2_{3;z2} = .2,\ R^2_{3;z23} = .4$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
57b	.2	-.2	0	.135	.192	.213	.120	.181	.203	.154	.177	.225	.142	.167	.215	.138	.231	.336
57c	.2	-.2	0	.125	.179	.202	.126	.183	.203	.142	.163	.212	.143	.168	.211	.132	.229	.327
58b	.2	-.2	-.2	.113	.182	.197	.109	.176	.186	.076	.085	.121	.073	.083	.113	.076	.154	.238
58c	.2	-.2	-.2	.106	.172	.184	.109	.177	.189	.073	.081	.113	.076	.086	.116	.072	.148	.231
59b	.2	-.2	.2	.139	.182	.217	.130	.171	.207	.390	.414	.483	.383	.408	.478	.338	.433	.553
59c	.2	-.2	.2	.130	.172	.204	.135	.176	.208	.382	.405	.474	.380	.405	.475	.335	.426	.546
60b	.2	-.2	-.5	.078	.156	.172	.079	.158	.173	.057	.051	.071	.057	.052	.075	.044	.108	.177
60c	.2	-.2	-.5	.072	.149	.165	.071	.146	.163	.056	.050	.072	.055	.050	.074	.045	.101	.170
61b	.2	-.2	.5	.118	.150	.208	.116	.145	.202	1	1	1	1	1	1	.999	1	1
61c	.2	-.2	.5	.111	.142	.201	.125	.156	.212	1	1	1	1	1	1	.999	1	1
62b	.2	.2	0	.120	.174	.195	.126	.181	.199	.139	.163	.208	.143	.168	.215	.124	.218	.329
62c	.2	.2	0	.131	.185	.207	.126	.182	.201	.149	.173	.221	.143	.166	.214	.132	.234	.332
63b	.2	.2	-.2	.124	.164	.196	.135	.178	.208	.376	.400	.471	.384	.409	.477	.330	.426	.544
63c	.2	.2	-.2	.134	.177	.214	.131	.175	.208	.391	.417	.485	.378	.402	.473	.342	.431	.547
64b	.2	.2	.2	.102	.171	.183	.104	.173	.187	.072	.080	.111	.073	.081	.115	.065	.145	.227
64c	.2	.2	.2	.112	.176	.192	.110	.176	.187	.075	.086	.117	.074	.084	.114	.068	.154	.239
65b	.2	.2	-.5	.111	.141	.194	.127	.155	.210	1	1	1	1	1	1	.999	1	1
65c	.2	.2	-.5	.120	.148	.212	.114	.145	.203	1	1	1	1	1	1	.999	1	1
66b	.2	.2	.5	.069	.145	.162	.070	.146	.162	.054	.048	.070	.053	.046	.070	.041	.096	.164
66c	.2	.2	.5	.075	.159	.173	.078	.158	.173	.059	.052	.077	.056	.049	.073	.045	.103	.177
67b	.5	-.2	0	.137	.210	.216	.811	.857	.879	.194	.221	.272	.848	.867	.899	.773	.889	.936
67c	.5	-.2	0	.127	.196	.205	.808	.860	.884	.181	.209	.255	.853	.872	.902	.773	.884	.934
68b	.5	-.2	-.2	.094	.202	.186	.781	.853	.875	.045	.052	.078	.733	.750	.802	.586	.766	.848
68c	.5	-.2	-.2	.091	.191	.174	.775	.847	.870	.050	.057	.082	.735	.753	.806	.590	.761	.844
69b	.5	-.2	.2	.159	.195	.229	.746	.767	.830	.882	.891	.919	.996	.997	.998	.993	.999	.999
69c	.5	-.2	.2	.147	.184	.223	.756	.777	.842	.878	.888	.918	.997	.997	.999	.992	.999	.999
70b	.5	-.2	-.5	.047	.172	.141	.598	.686	.769	.995	.995	.997	1	1	1	.987	1	1
70c	.5	-.2	-.5	.045	.163	.131	.592	.680	.758	.997	.995	.997	1	1	1	.988	1	1
71b	.5	.2	0	.119	.192	.198	.813	.861	.885	.178	.203	.256	.845	.866	.900	.771	.883	.928
71c	.5	.2	0	.131	.202	.210	.806	.855	.880	.189	.216	.270	.846	.866	.897	.775	.885	.935
72b	.5	.2	-.2	.142	.178	.215	.756	.778	.845	.876	.884	.914	.997	.998	.999	.992	.998	.999
72c	.5	.2	-.2	.158	.193	.227	.747	.767	.830	.882	.892	.920	.997	.997	.998	.993	.998	.999
73b	.5	.2	.2	.086	.190	.171	.777	.848	.871	.052	.058	.084	.735	.753	.806	.585	.761	.847
73c	.5	.2	.2	.094	.196	.177	.779	.853	.873	.049	.056	.081	.734	.753	.802	.586	.762	.846
74b	.5	.2	.5	.042	.159	.130	.588	.679	.759	.997	.996	.998	1	1	1	.988	1	1
74c	.5	.2	.5	.046	.170	.142	.599	.690	.768	.996	.995	.997	1	1	1	.987	1	1
75b	.5	-.5	0	.805	.838	.870	.794	.831	.862	.936	.947	.963	.939	.948	.963	.986	1	1
75c	.5	-.5	0	.800	.831	.862	.802	.836	.873	.937	.947	.962	.940	.949	.966	.987	1	1
76b	.5	-.5	-.2	.803	.893	.895	.794	.886	.887	.416	.439	.504	.404	.427	.496	.849	.972	.985
76c	.5	-.5	-.2	.799	.889	.889	.795	.888	.892	.408	.431	.496	.401	.425	.492	.849	.970	.984
77b	.5	-.5	-.5	.603	.804	.820	.599	.803	.816	.061	.055	.078	.061	.054	.081	.342	.683	.783
77c	.5	-.5	-.5	.598	.795	.809	.595	.793	.807	.057	.052	.078	.061	.054	.079	.340	.677	.774
78b	.5	.5	0	.800	.835	.865	.802	.838	.872	.933	.944	.959	.939	.948	.963	.985	1	1
78c	.5	.5	0	.809	.843	.872	.793	.828	.861	.935	.944	.961	.936	.946	.961	.987	1	1
79b	.5	.5	.2	.795	.889	.889	.794	.886	.890	.408	.434	.503	.414	.435	.501	.842	.971	.986
79c	.5	.5	.2	.804	.892	.892	.793	.887	.887	.416	.440	.507	.407	.431	.496	.844	.975	.987
80b	.5	.5	.5	.596	.794	.807	.593	.792	.808	.057	.051	.077	.059	.054	.077	.340	.677	.776
80c	.5	.5	.5	.603	.806	.816	.601	.799	.813	.060	.052	.078	.060	.054	.077	.346	.684	.783

Table 8: Two endogenous regressors and stronger instruments:
$R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
81b	.2	.2	0	.323	.350	.384	.331	.362	.394	.332	.327	.391	.338	.333	.400	.479	.497	.618
81c	.2	.2	0	.334	.361	.393	.330	.359	.391	.337	.332	.399	.336	.331	.397	.481	.502	.619
82b	.2	.2	-.2	.305	.314	.368	.318	.326	.380	.934	.928	.948	.937	.932	.951	.951	.948	.970
82c	.2	.2	-.2	.319	.328	.383	.317	.324	.378	.935	.930	.949	.936	.931	.952	.954	.950	.974
83b	.2	.2	.2	.290	.333	.358	.293	.339	.363	.100	.093	.129	.106	.098	.133	.232	.272	.384
83c	.2	.2	.2	.296	.343	.367	.296	.343	.368	.105	.098	.137	.105	.097	.133	.244	.282	.390
84b	.5	.2	0	.327	.360	.389	.999	.999	.999	.402	.395	.475	1	1	1	1	1	1
84c	.5	.2	0	.342	.375	.404	.999	.999	1	.408	.403	.483	1	1	1	1	1	1
85b	.5	.2	.2	.273	.357	.349	.999	.999	1	.200	.189	.250	.998	.998	.999	.999	1	1
85c	.5	.2	.2	.283	.366	.360	.999	.999	.999	.192	.179	.240	.997	.997	.998	.999	.999	1
86b	.5	.5	.2	1	1	1	1	1	1	.635	.613	.699	.640	.622	.702	1	1	1
86c	.5	.5	.2	.999	.999	1	1	1	1	.637	.619	.700	.631	.610	.692	1	1	1

Table 9: Two endogenous regressors and weak instruments for $y^{(3)}$:
$R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .01,\ R^2_{3;z23} = .02$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
87b	.2	.2	0	0	.060	.058	.105	.147	.147	.004	.162	.206	.338	.335	.398	.071	.256	.361
87c	.2	.2	0	.001	.063	.060	.105	.146	.146	.005	.161	.203	.340	.334	.398	.072	.258	.364
88b	.2	.2	-.2	0	.060	.060	.107	.136	.157	.089	.380	.443	.547	.537	.602	.158	.435	.556
88c	.2	.2	-.2	.001	.064	.065	.105	.131	.158	.092	.390	.450	.546	.535	.605	.157	.438	.553
89b	.2	.2	.2	.001	.061	.059	.045	.099	.077	.032	.172	.218	.259	.252	.311	.051	.194	.285
89c	.2	.2	.2	.001	.064	.061	.047	.102	.081	.030	.169	.213	.258	.250	.309	.050	.193	.285
90b	.2	.2	-.5	.001	.062	.074	.042	.078	.195	1	1	1	1	1	1	.735	1	1
90c	.2	.2	-.5	.001	.067	.084	.041	.075	.182	1	1	1	1	1	1	.731	1	1
91b	.2	.2	.5	.001	.060	.062	.006	.071	.038	.272	.280	.340	.311	.293	.353	.074	.224	.326
91c	.2	.2	.5	.001	.063	.066	.005	.069	.040	.270	.276	.336	.307	.290	.349	.074	.226	.327
92b	.5	.2	0	.001	.065	.074	.584	.466	.655	.075	.613	.644	1	1	1	.676	1	1
92c	.5	.2	0	.001	.069	.078	.576	.463	.643	.074	.612	.640	1	1	1	.669	1	1
93b	.5	.2	-.2	0	.068	.083	.449	.304	.572	.477	.900	.910	1	1	1	.731	1	1
93c	.5	.2	-.2	.001	.070	.092	.451	.295	.570	.483	.906	.916	1	1	1	.737	1	1
94b	.5	.2	.2	.001	.066	.077	.407	.317	.526	.388	.829	.851	1	1	1	.668	.999	1
94c	.5	.2	.2	.001	.068	.076	.404	.316	.524	.384	.828	.850	.999	.999	1	.662	.999	1
95b	.5	.5	0	.001	.093	.110	.615	.477	.682	.102	.679	.706	1	1	1	.702	1	1
95c	.5	.5	0	.002	.101	.119	.605	.473	.671	.100	.675	.699	1	1	1	.695	1	1
96b	.5	.5	.2	.001	.087	.097	.416	.361	.541	.356	.788	.812	.999	.999	.999	.634	.997	.999
96c	.5	.5	.2	.002	.098	.106	.413	.359	.534	.348	.790	.815	.999	.999	.999	.622	.997	.999
97b	.5	.5	.5	.001	.088	.102	.068	.156	.273	.998	.998	.999	1	.999	1	.676	.999	.999
97c	.5	.5	.5	.001	.092	.104	.071	.168	.285	.997	.997	.998	.999	.999	1	.670	.998	1

Table 10: Bootstrapped: One endogenous regressor and mildly strong instruments: $R^2_{2;z2} = .2,\ R^2_{2;z23} = .4,\ R^2_{3;z2} = .2,\ R^2_{3;z23} = .4$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
1b	0	0	0	.053	.048	.051	.050	.046	.048	.051	.051	.051	.047	.047	.047	.049	.050	.050
1c	0	0	0	.055	.052	.055	.051	.047	.050	.051	.051	.051	.049	.049	.048	.049	.050	.050
2b	0	0	-.2	.053	.051	.052	.052	.049	.050	.052	.052	.052	.049	.049	.049	.050	.050	.050
2c	0	0	-.2	.054	.051	.053	.051	.047	.049	.052	.052	.052	.046	.046	.046	.050	.051	.051
3b	0	0	.2	.052	.050	.052	.049	.046	.048	.052	.052	.052	.048	.048	.048	.048	.048	.048
3c	0	0	.2	.056	.052	.054	.050	.048	.049	.054	.054	.054	.046	.046	.046	.049	.049	.049
4b	0	0	-.5	.053	.051	.052	.049	.049	.050	.053	.053	.053	.054	.054	.054	.051	.049	.049
4c	0	0	-.5	.049	.049	.048	.047	.048	.047	.053	.053	.053	.051	.051	.051	.054	.052	.052
5b	0	0	.5	.052	.050	.052	.048	.048	.048	.050	.050	.050	.049	.049	.049	.047	.050	.050
5c	0	0	.5	.050	.049	.050	.049	.047	.048	.051	.051	.051	.047	.047	.047	.051	.050	.050
6b	.2	0	0	.054	.049	.051	.176	.164	.172	.051	.051	.051	.162	.162	.162	.125	.132	.132
6c	.2	0	0	.056	.052	.054	.177	.168	.174	.051	.051	.051	.159	.159	.159	.130	.138	.138
7b	.2	0	-.2	.052	.051	.051	.173	.162	.169	.079	.079	.079	.191	.191	.191	.150	.153	.153
7c	.2	0	-.2	.056	.052	.053	.172	.159	.170	.088	.088	.088	.192	.192	.192	.156	.156	.156
8b	.2	0	.2	.053	.051	.052	.169	.155	.166	.088	.088	.088	.195	.195	.195	.158	.160	.160
8c	.2	0	.2	.053	.052	.053	.171	.159	.169	.083	.083	.083	.190	.190	.190	.155	.156	.156
9b	.2	0	-.5	.051	.051	.052	.159	.138	.159	.680	.680	.680	.740	.740	.740	.697	.652	.652
9c	.2	0	-.5	.051	.048	.049	.147	.126	.148	.684	.684	.684	.742	.742	.743	.699	.650	.650
10b	.2	0	.5	.051	.051	.052	.150	.128	.150	.692	.692	.692	.750	.750	.750	.710	.657	.657
10c	.2	0	.5	.049	.049	.049	.162	.138	.161	.687	.687	.687	.743	.743	.743	.705	.654	.654
11b	.5	0	0	.050	.051	.049	.864	.847	.861	.053	.053	.053	.844	.844	.844	.775	.788	.788
11c	.5	0	0	.055	.055	.053	.861	.845	.858	.054	.054	.054	.842	.842	.842	.774	.786	.786
12b	.5	0	-.2	.050	.053	.048	.833	.788	.833	.342	.342	.342	.932	.932	.932	.896	.896	.896
12c	.5	0	-.2	.052	.053	.051	.823	.781	.824	.354	.354	.354	.932	.932	.932	.899	.900	.900
13b	.5	0	.2	.051	.054	.050	.827	.777	.825	.356	.356	.356	.938	.938	.938	.902	.901	.901
13c	.5	0	.2	.052	.053	.050	.833	.785	.832	.344	.344	.344	.935	.935	.935	.901	.901	.901

Table 11: Bootstrapped: One endogenous regressor and stronger instruments: $R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
14b	0	0	0	.053	.051	.052	0.047	.045	.046	.052	.052	.052	.048	.048	.048	.049	.049	.049
14c	0	0	0	.057	.054	.056	0.050	.049	.049	.053	.053	.053	.048	.048	.048	.050	.051	.051
15b	0	0	-.2	.051	.051	.051	0.051	.050	.050	.052	.052	.052	.051	.051	.051	.050	.050	.050
15c	0	0	-.2	.052	.050	.052	0.050	.049	.049	.054	.054	.054	.048	.048	.048	.051	.051	.051
16b	0	0	.2	.053	.053	.053	0.049	.049	.048	.050	.050	.050	.048	.048	.048	.048	.049	.049
16c	0	0	.2	.053	.051	.053	0.049	.049	.049	.055	.055	.055	.046	.046	.046	.049	.049	.049
17b	.2	0	0	.052	.052	.051	0.332	.326	.329	.047	.047	.047	.324	.324	.324	.249	.254	.254
17c	.2	0	0	.057	.056	.056	0.333	.328	.331	.047	.047	.047	.323	.323	.323	.255	.258	.258
18b	.2	0	.2	.052	.052	.051	0.313	.299	.312	.214	.214	.214	.465	.465	.466	.389	.378	.378
18c	.2	0	.2	.051	.050	.050	0.323	.308	.320	.208	.208	.208	.464	.464	.464	.386	.376	.376
19b	.2	0	-.2	.051	.051	.050	0.320	.306	.319	.205	.205	.205	.465	.465	.465	.389	.382	.382
19c	.2	0	-.2	.053	.052	.053	0.315	.298	.312	.208	.208	.208	.468	.468	.468	.387	.379	.379
20b	.5	0	0	.050	.052	.050	1	1	1	.025	.025	.025	1	1	1	.999	.999	.999
20c	.5	0	0	.055	.055	.054	1	1	1	.025	.025	.025	1	1	1	.999	.999	.999
21b	.5	0	.2	.049	.049	.050	.995	.989	.995	.976	.976	.976	1	1	1	1	1	1
21c	.5	0	.2	.049	.049	.048	.995	.989	.994	.976	.976	.976	1	1	1	1	1	1
22b	.5	0	-.2	.050	.053	.050	.994	.989	.994	.974	.974	.974	1	1	1	1	1	1
22c	.5	0	-.2	.049	.051	.049	.994	.989	.994	.975	.975	.975	1	1	1	1	1	1

Table 12: Bootstrapped: One endogenous regressor and weak instruments for $y^{(2)}$: $R^2_{2;z2} = .01,\ R^2_{2;z23} = .02,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
23b	0	0	0	.034	.038	.033	.029	.046	.042	.051	.051	.051	.050	.050	.050	.050	.049	.049
23c	0	0	0	.037	.038	.036	.027	.047	.042	.053	.054	.054	.048	.048	.048	.053	.050	.050
24b	0	0	.5	.014	.046	.024	.029	.047	.045	.051	.051	.051	.050	.050	.050	.050	.049	.049
24c	0	0	.5	.016	.049	.026	.028	.045	.042	.055	.055	.055	.054	.054	.054	.054	.051	.051
25b	.2	0	0	.033	.036	.032	.029	.052	.049	.049	.049	.049	.049	.049	.049	.049	.052	.052
25c	.2	0	0	.034	.038	.035	.033	.051	.047	.054	.054	.054	.052	.052	.052	.051	.052	.052
26b	.2	0	.5	.016	.046	.032	.030	.052	.048	.323	.322	.323	.325	.325	.325	.261	.257	.257
26c	.2	0	.5	.018	.049	.033	.033	.052	.050	.316	.316	.316	.315	.315	.315	.259	.248	.248
27b	.5	0	0	.034	.044	.034	.047	.079	.069	.049	.049	.049	.064	.064	.064	.046	.068	.068
27c	.5	0	0	.034	.044	.034	.046	.079	.071	.050	.050	.050	.060	.060	.060	.050	.069	.069
28b	.5	0	.5	.031	.048	.076	.046	.077	.077	1	1	1	1	1	1	.848	1	1
28c	.5	0	.5	.033	.050	.072	.048	.077	.083	1	1	1	1	1	1	.847	1	1

Table 13: Bootstrapped: One endogenous regressor and weak instruments for $y^{(3)}$: $R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .01,\ R^2_{3;z23} = .02$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
29b	0	0	0	.029	.050	.045	.034	.039	.033	.047	.047	.047	.049	.049	.049	.047	.048	.048
29c	0	0	0	.031	.050	.046	.033	.035	.032	.050	.050	.050	.049	.049	.049	.048	.050	.050
30b	0	0	.5	.033	.051	.048	.013	.051	.024	.047	.047	.047	.050	.050	.050	.048	.049	.049
30c	0	0	.5	.032	.052	.050	.013	.052	.024	.049	.049	.049	.048	.048	.048	.047	.050	.050
31b	.2	0	0	.031	.051	.047	.177	.141	.178	.157	.157	.157	.320	.320	0.320	.252	.253	.253
31c	.2	0	0	.031	.053	.050	.186	.145	.186	.159	.159	.159	.323	.323	0.323	.257	.253	.253
32b	.2	0	.5	.030	.050	.051	.038	.062	.074	.867	.867	.867	.876	.876	0.876	.688	.813	.813
32c	.2	0	.5	.032	.051	.054	.041	.060	.079	.866	.866	.866	.879	.879	0.879	.692	.813	.813
33b	.5	0	0	.030	.050	.054	.592	.399	.616	.599	.599	.599	1	1	1	.844	.999	.999
33c	.5	0	0	.031	.052	.057	.590	.404	.613	.594	.594	.594	1	1	1	.842	.999	.999
34b	.5	0	.2	.030	.052	.052	.388	.230	.451	.871	.871	.871	1	1	1	.861	1	1
34c	.5	0	.2	.030	.053	.059	.392	.239	.455	.866	.866	.866	1	1	1	.861	1	1

Table 14: Bootstrapped: One endogenous regressor and weak instruments: $R^2_{2;z2} = R^2_{3;z2} = .01,\ R^2_{2;z23} = R^2_{3;z23} = .02$

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	W_3^2	D_3^2	T_3^2	W^{23}	D^{23}	T^{23}
35b	0	0	0	.030	.038	.029	.028	.038	.028	.047	.047	.047	.048	.048	.048	.046	.047	.047
35c	0	0	0	.030	.040	.030	.029	.036	.029	.049	.049	.049	.048	.048	.048	.051	.049	.049
36b	0	0	.5	.031	.040	.032	.027	.043	.028	.048	.048	.048	.046	.046	.046	.045	.049	.049
36c	0	0	.5	.029	.036	.030	.026	.038	.027	.049	.049	.049	.049	.049	.049	.047	.049	.049
37b	.2	0	0	.029	.038	.028	.026	.037	.027	.051	.051	.051	.049	.049	.049	.048	.051	.051
37c	.2	0	0	.026	.039	.028	.030	.037	.029	.050	.050	.050	.050	.050	.050	.051	.053	.053
38b	.2	0	.5	.027	.042	.031	.028	.040	.030	.050	.050	.050	.053	.053	.053	.048	.056	.056
38c	.2	0	.5	.030	.043	.030	.026	.040	.028	.055	.055	.055	.056	.056	.056	.052	.054	.054
39b	.5	0	0	.029	.042	.031	.033	.048	.036	.059	.059	.059	.065	.065	.065	.053	.065	.065
39c	.5	0	0	.032	.043	.032	.034	.046	.036	.059	.059	.059	.060	.060	.060	.056	.067	.067
40b	.5	0	.5	.035	.057	.040	.045	.058	.047	.087	.087	.087	.094	.094	.094	.072	.100	.100
40c	.5	0	.5	.035	.054	.039	.040	.062	.047	.086	.086	.086	.092	.092	.092	.073	.098	.098

Table 15: Bootstrapped: Two endogenous regressors and mildly strong instruments: $R^2_{2;z2} = .2,\ R^2_{2;z23} = .4,\ R^2_{3;z2} = .2,\ R^2_{3;z23} = .4$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
57b	.2	-.2	0	.186	.175	.182	.179	.169	.175	.183	.183	.183	.167	.167	.167	.229	.242	.242
57c	.2	-.2	0.	.175	.166	.171	.179	.168	.175	.169	.169	.169	.170	.170	.170	.227	.239	.239
58b	.2	-.2	-.2	.172	.168	.169	.159	.160	.157	.087	.087	.087	.084	.084	.084	.138	.164	.164
58c	.2	-.2	-.2	.159	.160	.157	.162	.162	.160	.083	.083	.083	.085	.085	.085	.135	.158	.158
59b	.2	-.2	.2	.191	.164	.189	.183	.155	.179	.414	.414	.414	.408	.408	.408	.469	.447	.447
59c	.2	-.2	.2	.179	.154	.176	.183	.159	.179	.405	.405	.405	.408	.408	.408	.453	.438	.438
60b	.2	-.2	-.5	.141	.141	.140	.139	.142	.139	.052	.052	.052	.052	.052	.052	.084	.117	.117
60c	.2	-.2	-.5	.132	.132	.131	.129	.131	.131	.053	.053	.053	.051	.051	.052	.082	.108	.108
61b	.2	-.2	.5	.173	.127	.176	.165	.123	.168	1	1	1	1	1	1	1	1	1
61c	.2	-.2	.5	.163	.120	.162	.176	.134	.177	1	1	1	1	1	1	1	1	1
62b	.2	.2	0	.171	.160	.167	.176	.167	.173	.165	.165	.165	.169	.169	.169	.220	.233	.233
62c	.2	.2	0	.182	.172	.179	.176	.168	.173	.174	.174	.174	.165	.165	.165	.232	.244	.244
63b	.2	.2	-.2	.171	.149	.169	.184	.160	.180	.397	.397	.397	.409	.409	.409	.456	.435	.435
63c	.2	.2	-.2	.188	.161	.185	.184	.157	.181	.414	.414	.414	.399	.399	.399	.463	.440	.440
64b	.2	.2	.2	.158	.158	.155	.161	.160	.158	.082	.082	.082	.083	.083	.083	.128	.152	.152
64c	.2	.2	.2	.166	.165	.164	.164	.163	.161	.086	.086	.086	.083	.083	.083	.138	.164	.164
65b	.2	.2	-.5	.161	.119	.163	.175	.135	.179	1	1	1	1	1	1	1	1	1
65c	.2	.2	-.5	.176	.128	.178	.167	.123	.170	1	1	1	1	1	1	1	1	1
66b	.2	.2	.5	.128	.130	.128	.132	.133	.132	.047	.047	.047	.048	.048	.048	.076	.102	.102
66c	.2	.2	.5	.141	.140	.141	.141	.143	.141	.053	.053	.053	.051	.051	.051	.084	.111	.111
67b	.5	-.2	0	.198	.183	.193	.861	.835	.859	.219	.219	.219	.861	.861	.861	.883	.891	.891
67c	.5	-.2	0	.185	.170	.179	.863	.836	.861	.209	.209	.209	.866	.866	.866	.880	.886	.886
68b	.5	-.2	-.2	.158	.179	.156	.847	.833	.845	.054	.054	.054	.750	.750	.750	.732	.768	.768
68c	.5	-.2	-.2	.147	.167	.145	.842	.830	.842	.058	.058	.058	.750	.750	.750	.737	.771	.771
69b	.5	-.2	.2	.212	.165	.208	.806	.722	.806	.890	.890	.890	.997	.997	.997	.998	.999	.999
69c	.5	-.2	.2	.205	.153	.200	.815	.729	.817	.883	.883	.883	.997	.997	.997	.998	.998	.998
70b	.5	-.2	-.5	.104	.150	.107	.714	.654	.722	.994	.994	.994	1	1	1	.997	1	1
70c	.5	-.2	-.5	.096	.144	.099	.705	.644	.716	.995	.995	.995	1	1	1	.997	1	1
71b	.5	.2	0	.181	.167	.176	.870	.842	.866	.205	.205	.205	.864	.864	.864	.874	.883	.883
71c	.5	.2	0	.192	.177	.187	.860	.833	.857	.215	.215	.215	.861	.861	.861	.878	.889	.889
72b	.5	.2	-.2	.199	.148	.195	.820	.734	.822	.882	.882	.882	.997	.997	.997	.998	.998	.998
72c	.5	.2	-.2	.211	.163	.206	.809	.722	.808	.889	.889	.889	.997	.997	.997	.998	.998	.998
73b	.5	.2	.2	.147	.166	.145	.842	.828	.840	.060	.060	.060	.749	.749	.749	.734	.769	.769
73c	.5	.2	.2	.155	.174	.153	.846	.833	.845	.058	.058	.058	.746	.746	.746	.736	.767	.767
74b	.5	.2	.5	.097	.141	.098	.708	.644	.716	.995	.995	.995	1	1	1	.997	1	1
74c	.5	.2	.5	.102	.149	.106	.718	.651	.727	.994	.994	.994	.999	.999	.999	.997	1	1
75b	.5	-.5	0	.867	.793	.862	.859	.783	.854	.944	.944	.944	.946	.946	.946	.998	1	1
75c	.5	-.5	0	.856	.785	.854	.864	.791	.861	.945	.945	.945	.949	.949	.949	.998	1	1
76b	.5	-.5	-.2	.875	.873	.874	.863	.860	.862	.440	.440	.440	.427	.427	.427	.945	.972	.972
76c	.5	-.5	-.2	.864	.864	.863	.868	.862	.867	.430	.430	.430	.422	.422	.422	.940	.969	.969
77b	.5	-.5	-.5	.769	.779	.771	.765	.776	.766	.054	.054	.055	.054	.054	.054	.506	.693	.693
77c	.5	-.5	-.5	.759	.769	.762	.757	.766	.757	.052	.052	.052	.055	.055	.055	.501	.685	.685
78b	.5	.5	0	.862	.789	.858	.866	.789	.861	.940	.940	.940	.945	.945	.945	.998	1	1
78c	.5	.5	0	.867	.797	.865	.856	.781	.852	.944	.944	.944	.944	.944	.944	.998	1	1
79b	.5	.5	.2	.865	.864	.865	.867	.862	.867	.431	.431	.431	.432	.432	.432	.938	.970	.970
79c	.5	.5	.2	.872	.871	.871	.865	.861	.864	.440	.440	.440	.431	.431	.431	.944	.974	.974
80b	.5	.5	.5	.754	.767	.758	.756	.766	.760	.053	.053	.053	.055	.055	.055	.498	.685	.685
80c	.5	.5	.5	.765	.775	.767	.761	.771	.763	.054	.054	.054	.055	.055	.055	.510	.694	.694

Table 16: Bootstrapped: Two endogenous regressors and stronger instruments: $R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .3,\ R^2_{3;z23} = .6$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
81b	.2	.2	0	.331	.322	.329	.339	.332	.337	.324	.324	.324	.336	.336	.336	.506	.510	.510
81c	.2	.2	0	.340	.333	.338	.338	.331	.337	.330	.330	.330	.332	.332	.332	.508	.507	.507
82b	.2	.2	-.2	.318	.282	.316	.333	.297	.330	.924	.924	.924	.928	.928	.928	.954	.948	.948
82c	.2	.2	-.2	.330	.294	.329	.328	.292	.326	.927	.927	.927	.928	.928	.928	.958	.952	.952
83b	.2	.2	.2	.304	.308	.303	.308	.311	.306	.097	.097	.097	.099	.099	.099	.260	.283	.283
83c	.2	.2	.2	.308	.312	.308	.312	.315	.309	.102	.102	.102	.099	.099	.099	.272	.293	.293
84b	.5	.2	0	.345	.325	.343	.999	.998	.999	.395	.395	.395	1	1	1	1	1	1
84c	.5	.2	0	.357	.333	.354	.999	.999	.999	.403	.403	.403	1	1	1	1	1	1
85b	.5	.2	.2	.298	.324	.294	.999	.998	.999	.189	.189	.189	.998	.998	.998	.999	.999	.999
85c	.5	.2	.2	.307	.331	.305	.999	.998	.999	.181	.181	.181	.996	.996	.996	.999	.999	.999
86b	.5	.5	.2	.999	.999	.999	1	1	1	.609	.609	.608	.618	.618	.618	1	1	1
86c	.5	.5	.2	.999	.999	.999	1	1	1	.618	.618	.618	.603	.603	.603	1	1	1

Table 17: Bootstrapped: Two endogenous regressors and weak instruments for $y^{(3)}$: $R^2_{2;z2} = .3,\ R^2_{2;z23} = .6,\ R^2_{3;z2} = .01,\ R^2_{3;z23} = .02$

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	W^2_3	D^2_3	T^2_3	W^{23}	D^{23}	T^{23}
87b	.2	.2	0	.031	.053	.050	.187	.149	.187	.165	.165	.165	.334	.334	.334	.260	.268	.268
87c	.2	.2	0	.031	.055	.050	.186	.148	.189	.162	.162	.162	.335	.335	.335	.269	.266	.266
88b	.2	.2	-.2	.032	.052	.049	.159	.125	.171	.378	.378	.378	.535	.535	.535	.415	.447	.447
88c	.2	.2	-.2	.032	.056	.053	.158	.119	.168	.389	.389	.389	.534	.534	.534	.418	.446	.446
89b	.2	.2	.2	.031	.053	.049	.089	.097	.095	.175	.175	.175	.254	.254	.254	.204	.202	.202
89c	.2	.2	.2	.031	.056	.050	.095	.099	.101	.168	.169	.169	.252	.251	.251	.203	.203	.203
90b	.2	.2	-.5	.031	.051	.054	.056	.062	.127	1	1	1	1	1	1	.876	1	1
90c	.2	.2	-.5	.035	.056	.059	.051	.060	.117	1	1	1	1	1	1	.877	1	1
91b	.2	.2	.5	.034	.053	.048	.028	.062	.048	.281	.281	.281	.292	.292	.293	.237	.230	.230
91c	.2	.2	.5	.035	.055	.051	.027	.063	.049	.279	.279	.279	.288	.288	.289	.232	.236	.236
92b	.5	.2	0	.032	.055	.057	.599	.399	.626	.611	.611	.611	1	1	1	.846	1	1
92c	.5	.2	0	.032	.056	.058	.584	.401	.613	.611	.611	.611	1	1	1	.837	1	1
93b	.5	.2	-.2	.032	.051	.060	.404	.223	.487	.898	.898	.898	1	1	1	.877	1	1
93c	.5	.2	-.2	.033	.056	.067	.408	.215	.484	.906	.906	.906	1	1	1	.878	1	1
94b	.5	.2	.2	.031	.056	.058	.392	.250	.451	.829	.829	.829	.999	.999	.999	.843	.999	.999
94c	.5	.2	.2	.033	.057	.055	.389	.252	.450	.827	.827	.827	.999	.999	.999	.839	.999	.999
95b	.5	.5	0	.041	.076	.082	.613	.399	.648	.678	.678	.678	1	1	1	.856	1	1
95c	.5	.5	0	.046	.084	.089	.602	.397	.635	.674	.674	.674	1	1	1	.847	1	1
96b	.5	.5	.2	.042	.072	.073	.411	.296	.468	.788	.788	.788	.998	.998	.998	.811	.997	.997
96c	.5	.5	.2	.045	.082	.080	.407	.294	.464	.789	.789	.789	.999	.999	.999	.801	.997	.997
97b	.5	.5	.5	.044	.075	.076	.097	.132	.198	.998	.998	.998	.999	.999	.999	.837	.998	.998
97c	.5	.5	.5	.045	.078	.078	.102	.142	.213	.997	.997	.997	.999	.999	.999	.831	.998	.998

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	.0928	.0067	.1783	.0186	.1289	.0162	.1550	.0113
IQ	.0033	.0011	-.0099	.0052	-.0088	.0050	-.0017	.0013
EXPR	.0393	.0063	.0461	.0076	.0348	.0070	.0495	.0068
RNS	-.0745	.0288	-.1014	.0358	-.1096	.0341	-.0771	.0304
TEN	.0342	.0077	.0398	.0090	.0394	.0086	.0363	.0082
SMSA	.1367	.0279	.1291	.0321	.1475	.0305	.1212	.0296
CONS	3.8952	.1091	4.1049	.3552	4.6600	.3285	3.5641	.1244

Table 19: DWH tests for Griliches data

	Variables		Test Statistics			Critical values			
Test type	Tested	Instruments	W	D	T	$\chi^2_{.05}$	$\hat{W}^{bc}_{.05}$	$\hat{D}^{bc}_{.05}$	$\hat{T}^{bc}_{.05}$
Full-set	S, IQ	Z_1,Z_2	46.87	59.42	65.13	5.99	6.27	6.66	6.78
Full-set	S	Z_1, Z_2, IQ	50.56	55.90	60.96	3.84	4.69	4.70	4.77
Full-set	IQ	Z_1, Z_2, S	6.28	7.24	7.38	3.84	3.12	3.32	3.37
Sub-set	S	Z_1, Z_2	41.16	45.24	46.74	3.84	4.46	4.64	4.52
Sub-set	IQ	Z_1, Z_2	2.72	3.12	2.88	3.84	3.28	3.32	3.54