Discriminating between (in)valid external instruments and (in)valid exclusion restrictions
Kiviet, J.F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Discussion Paper: 2015/04

Discriminating between (in)valid external instruments and (in)valid exclusion restrictions

Jan F. Kiviet

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands
Discriminating between (in)valid external instruments and (in)valid exclusion restrictions

Jan F. Kiviet

Version of 26 November 2015
JEL Classifications: C01, C12, C26
Keywords: difference in Sargan-Hansen tests; exclusion restrictions; maintained hypothesis; instrument validity tests; omitted variables tests; overidentification restrictions

Abstract

In models estimated by (generalized) method of moments a test of coefficient restrictions can either be based on a Wald statistic or on the difference between evaluated criterion functions. Their correspondence can be used to demonstrate that a Sargan-Hansen test statistic for overidentification restrictions is equivalent to an omitted variables test statistic for a nonunique group of variables. We prove that this is the case for incremental Sargan-Hansen tests too. However, we also demonstrate that, despite this equivalence, one can nevertheless distinguish between either the (in)validity of some additional instruments or the (un)tenability of particular exclusion restrictions. It all hinges upon the required choice made regarding the initial maintained hypothesis.

1. Introduction

The classical trinity of test principles established in the context of Maximum Likelihood inference, constituted by Wald (W), Likelihood Ratio (LR) and Lagrange Multiplier (LM), have their counterparts in a semiparametric setting in which models can efficiently and robustly be estimated by (generalized) method of moments. In a linear method of moments context Section 2 focusses on correspondences between coefficient restriction tests based on either a standard W statistic or the test based on the difference between evaluated criterion functions (CF). The latter reminiscences of an LR-type test, but it can also be implemented as a kind of LM-type test. Section 3 shows that...
a particular implementation of a CF test gives the Sargan-Hansen statistic for testing overidentification restrictions (OR). The mandatory use of the latter test on a routine basis has been advocated at various places in the literature (see, for instance, examples in Baum et al., 2003, footnote 11), but warnings that it may mislead practitioners were recently issued in Parente and Santos Silva (2012) and Guggenberger (2012). We investigate the interpretation of a CF test as either a coefficient restriction test or as an OR instrument validity test and demonstrate that, although both use exactly the same test statistic, their respective maintained hypotheses differ. In Section 4 we prove that in linear models similar equivalences and differences exist between exclusion restrictions tests and so-called incremental overidentification restrictions (IOR) tests (also addressed as difference in Sargan-Hansen tests), which verify the validity of a subset of instruments. In the final Section 5 we conclude from our findings that strictly following formal statistical test principles enables to characterize the context in which OR and IOR tests should not mislead and in fact allow to distinguish inference regarding the (in)validity of some additional orthogonality conditions from the (un)tenability of particular exclusion restrictions. Practical problems do emerge though when one is unable or unwilling to condition the analysis on a firm initial maintained hypothesis.

2. Corresponding test principles

We focus on the single linear simultaneous regression model \( y = X\beta + \varepsilon \) with \( \varepsilon \sim (0, \sigma_\varepsilon^2 I) \), where \( X = (x_1, ..., x_n)' \) is an \( n \times K \) matrix of rank \( K \) with unknown coefficient vector \( \beta \in \mathbb{R}^K \). To cope with \( E(x_i\varepsilon_i) = \sigma_{x\varepsilon} \neq 0 \) for \( i = 1, ..., n \) we will employ an \( n \times L \) instrumental variables matrix \( Z = (z_1, ..., z_n)' \) of rank \( L \geq K \). The instrumental variables (IV) estimator is given by

\[
\hat{\beta} = (X'P_ZX)^{-1}X'P_Zy,
\]

where \( P_A = A(A'A)^{-1}A' \) for any full column rank matrix \( A \). When \( E(z_i\varepsilon_i) = 0 \) (the instruments are valid) and standard regularity conditions are fulfilled (including sufficient relevance of the instruments) then the IV estimator is consistent and asymptotically normal, whereas its variance can be estimated by \( \hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/n \) with residuals \( \hat{\varepsilon} = y - X\hat{\beta} \).

A test on any set of \( K_2 \) linear restrictions on the coefficients \( \beta \) can after a simple transformation of the model be represented as testing \( H_0: \beta_2 = 0 \) in

\[
y = X_1\beta_1 + X_2\beta_2 + \varepsilon \sim (0, \sigma_\varepsilon^2 I),
\]

where \( X_j \) is \( n \times K_j \) and \( \beta_j \) is \( K_j \times 1 \) for \( j = 1, 2 \). Making use of standard results on partitioned regression applied to the second-stage regression, in which \( y \) is regressed on
$P_ZX = (P_ZX_1, P_ZX_2) = (\hat{X}_1, \hat{X}_2)$, one easily finds that $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$, with

$$\hat{\beta}_2 = (\hat{X}_2'M_{X_1}\hat{X}_2)^{-1}\hat{X}_2'M_{X_1}y$$  \hspace{1cm} (2.3)

and $\widehat{Var}(\hat{\beta}_2) = \hat{\sigma}^2_e(\hat{X}_2'M_{X_1}\hat{X}_2)^{-1}$, where $M_A = I - P_A$. The W statistic for testing $H_0: \beta_2 = 0$ against alternative $H_1: \beta_2 \neq 0$ readily follows and is given by

$$W_{\beta_2} = \hat{\beta}_2'[\widehat{Var}(\hat{\beta}_2)]^{-1}\hat{\beta}_2.$$  \hspace{1cm} (2.4)

Under $H_0$ it is asymptotically $\chi^2(K_2)$ distributed, provided the maintained hypothesis does hold indeed. The maintained hypothesis (which is the union of $H_0$ and $H_1$ and all underlying assumptions) can here be characterized loosely as $M_1$: (i) the actual data generating process (DGP) is nested in model (2.2); (ii) $X$ and $Z$ show sufficient regularity; (iii) $n^{-1/2}Z'e \overset{d}{\rightarrow} N(0, \sigma^2_e \text{plim } n^{-1}Z'Z)$.

Substituting (2.3) and making use of a result on projection matrices for a partitioned full column rank matrix $A = (A_1, A_2)$ which says $P_A = P_{A_1} + P_{M_{A_1}A_2}$, we find

$$W_{\beta_2} = \gamma'P_{M_{X_1}X_2}y/\hat{\sigma}^2_e = \gamma'(P_X - P_{X_1})y/\hat{\sigma}^2_e = \gamma'(M_{X_1} - M_X)y/\hat{\sigma}^2_e.$$  \hspace{1cm} (2.5)

Hence, the test statistic can easily be obtained by taking the difference between the restricted and the unrestricted sum of squared residuals of second stage regressions, while scaling by a consistent estimator of $\sigma^2_e$. If the latter were obtained from the restricted residuals $\hat{\varepsilon} = y - X_1\hat{\beta}_1$, where

$$\hat{\beta}_1 = (\hat{X}_1'M_{X_1})^{-1}\hat{X}_1'y$$  \hspace{1cm} (2.6)

is the restricted IV estimator (imposing $\beta_2 = 0$) and $\hat{\sigma}^2_e = \hat{\varepsilon}'\hat{\varepsilon}/n$, the test statistic reminds of a Lagrange Multiplier test. Therefore we indicate it as

$$LM_{\beta_2} = \gamma'(M_{X_1} - M_X)y/\hat{\sigma}^2_e,$$  \hspace{1cm} (2.7)

which is asymptotically equivalent with $W_{\beta_2}$.

A test statistic with similarities to the LR principle is found as follows. Estimator $\hat{\beta}$ is obtained by minimizing with respect to $\beta$ a quadratic form in the vector $Z'(y - X\beta)$ in which a weighting matrix is used proportional to $(Z'Z)^{-1}$. This yields a consistent estimator with optimal variance under the maintained hypothesis. Defining the criterion function as

$$Q(\beta; y, X, Z, \sigma^2_e) = (y - X\beta)'Z(Z'Z)^{-1}Z'(y - X\beta)/\sigma^2_e,$$  \hspace{1cm} (2.8)

one can verify that it attains its minimum for $\hat{\beta}$ of (2.1), and for $\hat{\beta}_1$ of (2.6) in the model under $\beta_2 = 0$ upon using the criterion function $Q(\beta_1; y, X_1, Z, \sigma^2_e)$. Evaluating these criteria in their respective minima we find

$$Q(\hat{\beta}; y, X, Z, \sigma^2_e) = (y - X\hat{\beta})'P_Z(y - X\hat{\beta})/\sigma^2_e = (y'P_Zy - 2\hat{\beta}'X\hat{\beta} + \hat{\beta}'X'\hat{\beta})/\sigma^2_e$$

$$= (y'P_Zy - y'P_Xy)/\sigma^2_e = (y'M_Xy - y'Mzy)/\sigma^2_e.$$
with in the numerator the difference between the sum of the squares of the second-stage residuals and the reduced form residuals, and likewise

$$Q(\tilde{\beta}_1; y, X_1, Z, \sigma^2_\varepsilon) = (y'M_{\tilde{X}_1}y - y'M_{\tilde{X}_1}y)/\sigma^2_\varepsilon,$$

hence

$$Q(\tilde{\beta}_1; y, X_1, Z, \sigma^2_\varepsilon) - Q(\tilde{\beta}; y, X, Z, \sigma^2_\varepsilon) = (y'M_{\tilde{X}_1}y - y'M_{\tilde{X}_1}y)/\sigma^2_\varepsilon.$$

Thus, substituting either $\sigma^2_\varepsilon$ or $\tilde{\sigma}^2_\varepsilon$, we find that taking the difference between these two evaluated criterion functions yields a test statistic which has similarities with the log-likelihood-ratio principle, and since

$$CF_{\beta_2}(\tilde{\sigma}^2_\varepsilon) = Q(\tilde{\beta}_1; y, X_1, Z, \tilde{\sigma}^2_\varepsilon) - Q(\tilde{\beta}; y, X, Z, \tilde{\sigma}^2_\varepsilon) = W_{\beta_2},$$

(2.9)

is also algebraically equivalent to either the W or the LM statistic.

3. The Sargan test and redundant regressors

A special situation occurs in case $L = K$. Then the model is just identified, but overidentified under the $K_2$ restrictions $\beta_2 = 0$. When the model is just identified then in minimizing $Q(\beta; y, X, Z, \sigma^2_\varepsilon)$ the $L = K$ orthogonality conditions $Z'(y - X\hat{\beta}) = 0$ will all be satisfied in the sample, giving $Q(\hat{\beta}; y, X, Z, \sigma^2_\varepsilon) = 0$, so in this special situation (indicated by an asterisk) the maintained hypothesis is $M_1^*$ and

$$W_{\beta_2}^* = CF_{\beta_2}^*(\tilde{\sigma}^2_\varepsilon) = Q(\tilde{\beta}_1; y, X_1, Z, \tilde{\sigma}^2_\varepsilon) = \varepsilon'P_2\tilde{\varepsilon}/\tilde{\sigma}^2_\varepsilon,$$

(3.1)

whereas an asymptotically equivalent statistic is given by

$$LM_{\beta_2}^* = CF_{\beta_2}^*(\tilde{\sigma}^2_\varepsilon) = \varepsilon'P_2\tilde{\varepsilon}/\tilde{\sigma}^2_\varepsilon.$$

(3.2)

The latter test statistic is well-known. It is the Sargan test for the $L - K_1$ overidentifying restrictions in the model which has just the regressors $X_1$, say

$$y = X_1\beta_1^* + \varepsilon^*$$

with $\varepsilon^* \sim (0, \sigma^2_\varepsilon, I).$

(3.3)

From its algebraic equivalence with a coefficient restriction test it is immediately obvious that testing for the overidentification restrictions has similarities with testing the validity of $L - K_1$ exclusion restrictions.

However, there are dissimilarities as well. Validity of the $L$ orthogonality conditions $E(z_i\varepsilon_i) = 0$ is part of the maintained hypothesis $M_1^*$ for the test of $L - K_1$ exclusion restrictions, whereas the Sargan test is generally used to verify validity of the orthogonality conditions $E(z_i\varepsilon_i^*) = 0$. Although using exactly the same test statistic and critical
values, we will show below that the two tests have different \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) and build on different maintained hypotheses as well.

Note that expression (3.2) is invariant regarding \( X_2 \). So, replacing \( X_2 \) by any \( n \times K_2 \) matrix \( X_2^* \) would yield equivalent results for \( CF_{\beta_2}^* (\tilde{\sigma}_2^2) \), provided \( P_{\tilde{Z}}(X_1, X_2^*) \) has full column rank \( K = L \). Thus, the variables \( X_2^* \) should be sufficiently related to the instruments \( Z \), but at the same time they may also be endogenous regarding \( \varepsilon \). Due to the projection of the regressors on the space spanned by the instruments \( Z \) in the second stage regression, and the presence of the regressors \( X_1 \), the coefficient restriction test actually does not test the explanatory power of \( X_2^* \), but just that of \( M_{X_1} P_{\tilde{Z}} X_2^* = (P_{\tilde{Z}} - P_{\tilde{X}_1} P_{\tilde{Z}}) X_2^* = (P_{\tilde{Z}} - P_{\tilde{X}_1}) X_2^* \).

A viable choice for \( X_2^* \) would be the following. Let \( H = (H_1, H_2) \) be a full rank \( L \times L \) matrix, where \( H_1 \) is \( L \times K_1 \) and \( Z^* = (Z_1^*, Z_2^*) = ZH = (ZH_1, ZH_2) \) with \( H_1 \) such that \( Z_1^* X_1 \) has full rank. Obviously, \( H \) is nonunique. If the transformed instruments \( Z_1^* \) are not just relevant for \( X_1 \), but were valid for model (3.3) as well, then the coefficients \( \beta_1^* \) would be just identified by them. Now consider taking \( X_2^* = M_{Z_1^*} Z_2^* \). Then \( P_{\tilde{Z}}(X_1, X_2^*) \) has full column rank as required, and statistic \( CF_{\beta_2}^* (\tilde{\sigma}_2^2) \) for testing \( \mathcal{H}_0 : \beta_2^* = 0 \) in model
\[
y = X_1 \beta_1^* + M_{Z_1^*} Z_2^* \beta_2^* + \varepsilon^*
\]
is therefore equal to \( CF_{\beta_2}^* (\tilde{\sigma}_2^2) \). Estimating \( \beta_1^* \) of (3.4) by IV using the instruments \( Z \) or \( Z^* \) (which is equivalent because \( P_{\tilde{Z}} = P_{\tilde{Z}^*} ) \) yields
\[
\hat{\beta}_1^* = (X_1 P_{Z_1^*} X_1)^{-1} X_1 P_{Z_1^*} y = (Z_1^* X_1)^{-1} Z_1^* y = \hat{\beta}_1^*,
\]
because \( P_{Z^*} M_{Z_1^*} Z_2^* P_{Z^*} = P_{Z^*} - P_{M_{Z_1^*} Z_2^*} = P_{Z_1^*} \). Hence, augmenting model (3.3) by \( K_2 \) regressors \( M_{Z_1^*} Z_2^* \) and using instruments \( Z \) or \( Z^* \) yields coefficient estimates for the regressors \( X_1 \) equivalent to the simple IV estimator obtained in model (3.3) when just using the \( K_1 \) instruments \( Z_1^* \), irrespective of the validity of any of the regressors.

Apart from the interpretation of \( CF_{\beta_2}^* (\tilde{\sigma}_2^2) \) as testing \( K_2 = L - K_1 \) exclusion restrictions \( \beta_2 = 0 \) in model (2.2) under \( E(z_i \varepsilon_i) = 0 \), which is based on maintained hypothesis \( M_{Z_1^*}^2 \), yet another coherent interpretation of the numerically equivalent test statistic \( CF_{\beta_2}^* (\tilde{\sigma}_2^2) \) is the following. It tests whether the \( L - K_1 \) variables \( Z_2^* \) are additional valid external instruments for model (3.3) under the maintained hypothesis given by \( M_{Z_1^*}^2 \): (i) the DGP is nested in model (3.3); (ii) \( X_1 \) and \( Z \) (and thus \( Z^* \)) show sufficient regularity; (iii) \( n^{-1/2} Z_1^* \varepsilon^* \overset{d}{\rightarrow} N(0, \sigma_{\varepsilon^*}^2) \), \( \overline{\text{plim}} n^{-1} Z_1^* Z_1^* \).

Otherwise, if the variables \( Z_2^* \) (like \( Z_1^* \)) are uncorrelated with \( \varepsilon^* \) then \( M_{Z_1^*} Z_2^* \) should not have explanatory power in model (3.4). Note that in this alternative context all instruments \( Z \) are assumed to be valid for model (3.3) under the null hypothesis, but under the alternative \( E(z_2^* \varepsilon_2^*) = \sigma_{z_2^* \varepsilon_2^*} \neq 0 \). Hence, the null hypothesis is given by the
$L - K_1$ orthogonality conditions $E(z_{2i}^* \varepsilon_i^*) = 0$. When these are rejected and thus the estimates of model (3.4) using the instruments $Z^*$ or $Z$ are preferred, then, due to (3.5), one still obtains an estimator for $\beta_1^*$ which is consistent under the maintained hypothesis $M_2^*$. In regression (3.4) $Z_2^*$ constitutes valid instruments regarding $\varepsilon^*$ under $M_2^*$, because inclusion of the regressors $M_2 Z_1^* Z_2^*$ purges the disturbances from their correlation with $Z_2^*$.

It is the case that truth of the null $\beta_2 = 0$ under $M_1^*$ implies $E(z_i \varepsilon_i^*) = 0$, and so does truth of the null $E(z_{2i}^* \varepsilon_i^*) = 0$ under $M_2^*$. Hence, in both cases, imposing the respective null hypothesis leads to using estimator $\tilde{\beta}_1$ for $\beta_1$. However, this is due either to imposing coefficient restrictions or to imposing extra orthogonality restrictions. Rejection of $\beta_2 = 0$ under $M_1^*$ leads to preferring $\hat{\beta}$ for $\beta$, but rejection of $E(z_{2i}^* \varepsilon_i^*) = 0$ under $M_2^*$ to preferring $\tilde{\beta}_1$ for $\beta_1^*$ with no further regression coefficients, since in that setup the DGP is supposed to be nested in model (3.3). So, also in case of rejection, the actually adopted maintained hypothesis forces to discriminate, namely between either invalidity of $L - K_1$ exclusion restrictions or invalidity of $L - K_1$ external instruments.

The warnings of Parente and Santos Silva (2012) and Gugenberger (2012) concern the problems that emerge when using the Sargan statistic for testing jointly the $L$ orthogonality conditions $E(z_i \varepsilon_i^*) = 0$. However, using this $L$ dimensional null hypothesis, is asking too much from this $L - K_1$ degrees of freedom test statistic, leading to inconsistency of the test (existence of nonlocal alternatives to the null that are not rejected with probability one asymptotically). The classic Sargan test can only pretend to serve those well who are willing to adopt maintained hypothesis $M_2^*$, involving the statistically untestable orthogonality conditions expressed by $E(z_{1i}^* \varepsilon_i^*) = 0, i = 1, ..., n$. It is the researcher who may decide whether (s)he leaves these orthogonality conditions implicit, or makes them explicit by actually specifying the matrix $H_1$, for instance by choosing $H_1 = (I, O)'$ and then thus simply assuming that the first $K_1$ instruments in the matrix $Z$ are valid in model (3.3). In that special case the explicit null hypothesis of the Sargan test is that the instruments $M_2 Z_1 Z_2$ are valid too, which would imply $E(z_{i2} \varepsilon_i^*) = 0$.

4. The incremental Sargan test and redundant regressors

In the model considered above the validity of a subset of the instruments can be tested by a so-called incremental Sargan test. For this case too we will examine in what way such a procedure has (dis)similarities with testing exclusion restrictions. Instead of testing the validity of instruments in model (3.3) with $K_1 < K = L$ we will here take the model with $K$ regressors and $L$ instruments as our starting point.

We consider now a partition of the matrix of $L (> K)$ instruments, denoted as $Z = (Z_m, Z_a)$, where for $j \in \{m, a\}$ matrix $Z_j = (z_{j1}, ..., z_{jn})'$ is $n \times L_j$. Matrix $Z_m$, with
\( L_m \geq K \), contains the instruments which are maintained to be valid, whereas the validity of the additional \( L_a \) instruments \( Z_a \) will be examined. So the maintained hypothesis is now \( \mathcal{M}_3 \): (i) the DGP is nested in \( y = X\beta + \varepsilon \); (ii) \( X \) and \( Z \) show sufficient regularity; (iii) \( n^{-1/2}Z_m'\varepsilon \overset{d}{\rightarrow} \mathcal{N}(0, \sigma^2_{\varepsilon} \text{plim} n^{-1}Z_m'Z_m) \).

Just using the instruments \( Z_m \) yields estimator and residuals

\[
\hat{\beta}_m = (X'P_{Zm}X)^{-1}X'P_{Zm}y, \quad \hat{\varepsilon}_m = y - X\hat{\beta}_m.
\]

From the foregoing section it follows that when \( L_m > K \), the Sargan statistic

\[
S_m = n \cdot \hat{\varepsilon}'_m P_{Zm} \hat{\varepsilon}_m / \hat{\varepsilon}'_m \hat{\varepsilon}_m
\]

is asymptotically \( \chi^2(L_m - K) \) distributed under \( \mathcal{M}_3 \). Using the additional instruments as well yields \( \hat{\beta} \) and \( \hat{\varepsilon} \) and enables to calculate

\[
S_{ma} = n \cdot \hat{\varepsilon}' P_{Z} \hat{\varepsilon} / \hat{\varepsilon}' \hat{\varepsilon}.
\]

The null hypothesis \( E(z_{ia}\varepsilon_i) = 0 \) can now be tested by the incremental Sargan statistic

\[
IS_a = S_{ma} - S_m.
\]

When deriving its null distribution we will assume, without loss of generality, that \( Z_m'Z_a = O \). Supposing that we have replaced the original \( Z_a \) by \( Z_m'Z_a \), then \( P_Z \) is not affected but now \( P_Z = P_{Zm} + P_{Za} \). From \( P_Z \hat{\varepsilon} = (P_Z - P_{P_{Z}X})\varepsilon \) we find \( \hat{\varepsilon}'P_Z\hat{\varepsilon} = \varepsilon'P_Z\varepsilon - \varepsilon'P_{P_{Z}X}\varepsilon = \varepsilon'(P_{Zm} + P_{Za} - P_{P_{Z}X})\varepsilon \) and so \( \hat{\varepsilon}'P_Z\hat{\varepsilon} - \hat{\varepsilon}'_m P_{Zm} \hat{\varepsilon}_m = \varepsilon'(P_{Za} + P_{Zm}x - P_{P_{Z}X})\varepsilon \).

The matrix in this quadratic form is symmetric and idempotent. The latter follows from

\[
tr(P_{Za} + P_{Zm}x - P_{P_{Z}X}) = L_a + K - K = L_a
\]

this means that under \( \mathcal{M}_3 \) and the null hypothesis, so when all instruments are valid, \( IS_a \overset{d}{\rightarrow} \varepsilon'(P_{Za} + P_{P_{Z}X} - P_{P_{Z}X})\varepsilon / \sigma^2_{\varepsilon} \overset{d}{\rightarrow} \chi^2(L_a) \), because under the null both \( \text{plim} \hat{\varepsilon}'_m \hat{\varepsilon}_m / n = \sigma^2_{\varepsilon} \) and \( \text{plim} \hat{\varepsilon}' \hat{\varepsilon} / n = \sigma^2_{\varepsilon} \).

We shall now examine tests for omitted regressors and seek correspondences with \( IS_a \). Consider the extended model with \( K_a \) additional regressors \( X_a \) given by

\[
y = X\beta + X_a\beta_a + \varepsilon_a, \quad \text{with} \quad \varepsilon_a \sim (0, \sigma^2_{\varepsilon_a} I).
\]

If \( X_a \) is such that estimation by IV using instruments \( Z \) is possible and yields residuals \( \varepsilon_a \), then the LM-like test statistic for the exclusion restrictions \( \beta_a = 0 \) is

\[
CF_{\beta_a}(\hat{\sigma}^2_{\varepsilon}) = (\hat{\varepsilon}'P_Z\hat{\varepsilon} - \varepsilon_a'P_{Z}\varepsilon_a)/\hat{\sigma}^2_{\varepsilon}.
\]

which has null distribution \( \chi^2(K_a) \) under the appropriate maintained hypothesis.

Next we specialize to the specific case \( CF_{\beta_a}^*(\hat{\sigma}^2_{\varepsilon}) \), where \( X_a = Z_a \), and examine its properties to test \( H_0 : \beta_a^* = 0 \) under the maintained hypothesis \( \mathcal{M}_a^* \): (i) the DGP is
nested in model \( y = X\beta^* + Z_a\beta_a^* + \varepsilon_a^* \); (ii) \((X, Z_a)\) and \(Z\) show sufficient regularity; (iii) \( n^{-1/2}Z_d\varepsilon_a^* \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon_a}^2) \) plim \( n^{-1}Z'Z\). Indicating the IV results as \( y = X\beta^* + Z_a\beta_a^* + \varepsilon_a^* \), and (since we previously replaced \( Z_a \) by \( Z_m Z_a \)) using \( Z_m Z_a = O \), we find

\[
\hat{\beta}^* = (X'P_Z M_{Z_m} P_Z X)^{-1} X'P_Z M_{Z_m} y = (X'P_{Z_m} X)^{-1} X'P_{Z_m} y = \hat{\beta}_m,
\]

(4.7)

because \( M_{Z_m} P_Z = P_Z - P_{Z_m} P_Z = P_Z - P_{Z_m} = P_{Z_m} \).

Since in IV estimation the residuals are orthogonal to the second-stage regressors we have \( Z_a \varepsilon_a^* = 0 \). This implies that \( P_Z \varepsilon_a^* = (P_{Z_m} + P_{Z_a}) \varepsilon_a^* = P_{Z_m} (y - X\hat{\beta}^* - Z_a\hat{\beta}_a^*) = P_{Z_m} (y - X\hat{\beta}_m) = P_{Z_m} \varepsilon_m. \) Hence,

\[
CF_{\beta_a^*}^*(\hat{\sigma}_\varepsilon^2) = \frac{(\varepsilon' P_Z \hat{\varepsilon} - \varepsilon'_a P_Z \varepsilon_a^*)/\hat{\sigma}_\varepsilon^2}{(\varepsilon'_m P_{Z_m} \hat{\varepsilon}_m)/\hat{\sigma}_\varepsilon^2} = S_{ma} - S_m \frac{\varepsilon'_m \hat{\varepsilon}_m}{\varepsilon'_\varepsilon}.
\]

(4.8)

Under \( \mathcal{M}_1^* \) and \( H_0 : \beta_a^* = 0 \) we have \( E(z_i \varepsilon_i) = 0 \) so the ratio \( \varepsilon'_m \hat{\varepsilon}_m/\varepsilon'_\varepsilon \) equals 1 asymptotically. Therefore, under their respective null hypotheses, the incremental Sargan test statistic \( IS_a \) is asymptotically equivalent with the omitted variables test statistic \( CF_{\beta_a^*}^*(\hat{\sigma}_\varepsilon^2) \).

In fact, this also holds for more general matrices than just \( X_a = Z_a \). Consider instead

\[
X_a = Z_a C_1 + XC_2 + M_Z C_3 + \varepsilon C_4,
\]

(4.9)

where the finite matrices \( C_1, C_2, C_3, C_4 \) are \( L_a \times L_a, K \times L_a, n \times L_a \) and \( 1 \times L_a \) respectively. Now consider the relaxation of \( \mathcal{M}_1^* \) labelled \( \mathcal{M}_4 \): (i) the DGP is nested in model (4.5) with (4.9); (ii) \((X, X_a)\) and \(Z\) show sufficient regularity; (iii) \( n^{-1/2}Z_d\varepsilon_a^* \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon_a}^2) \) plim \( n^{-1}Z'Z\). Now writing \( \hat{X}_a = P_Z X_a \) and \( \hat{X} = P_Z X \) it follows that \( \hat{X}_a = Z_a C_1 + \hat{X} C_2 + P_Z \varepsilon C_4 \). Under the null hypothesis \( \beta_a = 0 \), we have plim \( n^{-1}Z'\varepsilon = 0 \), hence \( P_Z \varepsilon C_4 \rightarrow 0 \), thus \( M_{\hat{X}} \hat{X}_a \rightarrow M_{\hat{X}} Z_a C_1 \) and \( P_{M_{\hat{X}}} \hat{X}_a \rightarrow P_{M_{\hat{X}}} Z_a \). Since \( CF_{\beta_a^*}^*(\hat{\sigma}_\varepsilon^2) = y'P_{M_{\hat{X}}} \hat{X}_a y/\hat{\sigma}_\varepsilon^2 \) and \( CF_{\beta_a^*}^*(\hat{\sigma}_\varepsilon^2) = y'P_{M_{\hat{X}}} \varepsilon y/\hat{\sigma}_\varepsilon^2 \), tests for the omission of regressors \( X_a \) which belong to group (4.9) will yield test statistics which under the null are all asymptotically equivalent with statistic \( IS_a \).

If \( CF_{\beta_a^*}^*(\hat{\sigma}_\varepsilon^2) \) and \( IS_a \) do not reject and one wants to impose their null then estimator \( \hat{\beta} \) should be preferred, irrespective whether one had adopted \( \mathcal{M}_3 \) (and accepts the instruments \( Z_a \)) or \( \mathcal{M}_4 \) (and accepts exclusion restrictions \( \beta_a = 0 \)). If a rejection is found and one had adopted \( \mathcal{M}_3 \) then \( \hat{\beta}_m \) should be preferred (and thus the instruments \( Z_a \) skipped), whereas under starting point \( \mathcal{M}_4 \) the preferred estimators are obtained now from model (4.5). This yields an estimator for \( \beta \) equal to \( \hat{\beta}_m \) only if \( C_2 = O \) and \( C_4 = 0' \). After rejection the degree of overidentification decreases by \( L_a \) in comparison to imposing the respective nulls; this is either due to discarding the candidate instruments
or by not excluding the regressors $X_a$. Note that Section 3 deals in fact with the special case $L_a = L - K$, $L_m = 0$ and $S_m = 0$.

Result (4.7) reestablishes the so-called partialling out result, which says that IV coefficient estimates are invariant when (putative) exogenous regressors are deliberately omitted, provided they are also removed from and filtered out of the remaining instruments. In Hendry (2011) this analytic result is established by simulation, which leads to the unsatisfactory conclusion that $\hat{\beta}$ and $\hat{\beta}_m$ "hardly differ". Moreover, in the simulation design the instruments are all valid, $L = 3$, $K = 2$, $L_a = 1$ and all variables are normally distributed, whereas neither of these is required for the algebraic result to hold exactly.

5. Interpretation and conclusion

The above results lead to clear guidelines only when a researcher is willing to adopt unambiguously a particular maintained hypothesis, which (s)he should according to sound statistical methodology. In practice, however, a more flexible though opportunistic approach may often seem more appealing. The two juxtaposed maintained hypotheses in Section 3 (and their generalizations in Section 4) both suppose that knowledge is available already of a model specification in which the DGP is nested, whereas endeavors to reach that stage form usually the major challenge in an empirical modelling exercise. This explains why researchers (also see Cameron and Trivedi, 2005, p.277) when obtaining a significant OR test may either conclude: (a) some instruments are invalid, (b) the model specification is deficient (so its implied disturbances may correlate even with valid instruments for an adequate model specification), or (c) both model and instruments are unsound. We have shown that this shillyshelly is a consequence of not building on a firm maintained hypothesis, and therefore an insignificant OR statistic, apart from possibly due to a type I error, may occur due to either (a), (b) or (c) as well.

The presented results on instrument validity and coefficient restriction tests can rather straight-forwardly be generalized to models with nonspherical disturbances estimated by generalized method of moments. In that case one refers to Sargan-Hansen tests, often indicated as (differences) in $J$ tests. With some more effort generalizations for nonlinear models can be obtained too. Also in that context the practical relevance of these findings is the following. An (incremental) Sargan-Hansen test on the validity of (particular) orthogonality conditions uses exactly the same (or at least an under the null asymptotically equivalent) test statistic with the same asymptotic null distribution as a test for zero restrictions on the coefficients of additional regression variables which may stem from some fairly wide group. However, these two test procedures build on different nonnested maintained hypotheses which enables them to produce discriminative
If the chosen maintained hypothesis implies overidentification then by deeper implementations of the tests it is possible to verify the tenability of particular aspects of this maintained hypothesis, but not all of them. In the statistical framework for analyzing (partial) simultaneous equations models an initial (at least just-)identifying maintained hypothesis has to be adopted. To do this evidence is required from a different source than statistical inference obtained from the same data set. Next, by a sequence of tests on either possibly redundant regressors or additional valid instruments the initial maintained hypothesis could be further restrained to reach higher levels of overidentification. This sequence of tests can be designed such that separate tests form links in a chain, where the individual test statistics are asymptotically independent under the union of the preceding series of accepted and incorporated null hypotheses and the initial maintained hypothesis, see for instance Godfrey and Hutton (1994), which allows then to control the overall significance level.

References


