UvA ECONOMETRICS

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

UvA UNIVERSITEIT VAN AMSTERDAM

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

Jan F. Kiviet*

Version of 5 May 2016

JEL Classifications: *C01, C12, C26*

Keywords: *difference in Sargan-Hansen tests; exclusion restrictions; maintained hypothesis; instrument validity tests; overidentifying restrictions*

Abstract

In models estimated by (generalized) method of moments a test on coefficient restrictions can either be based on a Wald statistic or on the difference between evaluated criterion functions. From their correspondence it easily follows that the statistic used for testing instrument validity, the Sargan-Hansen overidentifying restrictions (OR) statistic, is equivalent to an exclusion restrictions test statistic for a nonunique group of regressor variables. We prove that asymptotically this is the case too for incremental OR tests. However, we also demonstrate that, despite this equivalence of test statistics, one can nevertheless distinguish between either the (in)validity of some additional instruments or the (un)tenability of particular exclusion restrictions. This, however, requires to be explicit about the adopted maintained hypothesis. It also highlights that recent warnings in the literature that overidentifying restrictions tests may mislead practitioners should not be directed towards the test, but to practitioners who do not realize that inference based on such tests is unavoidably conditional on the validity of particular just-identifying statistically untestable assumptions.

*Emeritus Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands (j.f.kiviet@uva.nl). Most of this paper was written while I was enjoying hospitality as Visting Professor at the Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332. Helpful comments by João Santos Silva on an earlier version of this paper and further suggestions by JEM co-editor Jason Abrevaya are gratefully acknowledged.

1. Introduction

In empirical econometric models it is often not unlikely that some explanatory variables are in fact endogenous, because their realizations are contemporaneously correlated with the error terms of the model. To overcome inference problems due to such correlation one often exploits external (non-explanatory) variables that should be uncorrelated with these errors. To verify whether these variables (instruments) seem uncorrelated with the disturbances indeed one can use a so-called test for overidentifying restrictions (OR), see Sargan (1958), Hansen (1982). The mandatory use of OR tests on a routine basis has been advocated at various places in the literature. See, for instance, the list mentioned in Baum et al. (2003, footnote 11). However, warnings that OR tests may mislead practitioners were recently issued in Parente and Santos Silva (2012) and Guggenberger (2012), and can also be found in, for instance, Hayashi (2000, p.218) and Cameron and Trivedi (2005, p.277).

In this study we will place these warnings regarding OR tests into perspective and highlight that they only apply if one is hesitant with respect to formulating an explicit maintained hypothesis, involving sufficient a priori orthogonality conditions and exclusion restrictions. Also we will expose the algebraic equivalences between OR and CR (coefficient restrictions) test statistics and explain why these tests nevertheless allow distinct inferences. Moreover, we discuss to what degree similar issues arise for incremental OR tests.

For the sake of simplicity we will focus here mainly on the linear method of moments context, but our derivations can be extended to more general settings, as we shall indicate. First, while introducing our notation, we review some general results on testing in a linear instrumental variables context, such as illustrating that 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, see also Newey and West (1987). In a linear method of moments context CR tests based on a W statistic are in fact equivalent to the test obtained by taking the difference between evaluated criterion functions (CF) for the restricted and the unrestricted model. The latter reminiscences of an LR-type test. By employing an estimate for the disturbance variance obtained from the restricted model both W and CF can be implemented as a kind of LM-type test. A very particular implementation of a CF test gives the Sargan-Hansen statistic for testing overidentifying restrictions. Thus, the very same test statistic can be interpreted either as a CR test or as an OR instrument validity test. However, we will highlight that these two interpretations require different maintained hypotheses. This has not always been made very clear in earlier literature.

We also prove that in linear models similar equivalences and differences exist between exclusion restrictions tests and so-called incremental overidentifying restrictions (IOR) tests (also addressed as difference in Sargan-Hansen tests), which verify the validity of a subset of instruments. From these findings we conclude 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 inference on the (un)tenability of particular exclusion restrictions. Practical problems do emerge only when one is unable or unwilling to condition the analysis on a firm initial maintained hypothesis.

The structure of this study is as follows. In Section 2 we introduce the model and various test procedures and concepts. Section 3 demonstrates how the standard Sargan test can either be used for testing instrument validity or under a different maintained hypothesis can be interpreted as a CR test. Section 4 demonstrates to what degree these results do apply to incremental Sargan tests as well. Finally Section 5 indicates options for further generalizations and summarizes the conclusions.

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) or two-stage least-squares (2SLS) estimator is given by

$$\hat{\beta} = (X'P_Z X)^{-1} X'P_Z y, \tag{2.1}$$

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 too (including sufficient relevance of the instruments) then the IV estimator is consistent and asymptotically normal, whereas its variance can be estimated by $\widehat{Var}(\hat{\beta}) = \hat{\sigma}_\varepsilon^2 (X'P_Z X)^{-1}$, where $\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/n$ with residuals $\hat{\varepsilon} = y - X\hat{\beta}$.

A test on any set of $K_2 < K$ linear restrictions on the coefficients β can after a simple transformation of the model be represented as testing exclusion restrictions $\mathcal{H}_0 : \beta_2 = 0$ in

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \text{ with } \varepsilon \sim (0, \sigma_\varepsilon^2 I), \tag{2.2}$$

where X_j is $n \times K_j$ and β_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

$P_Z X = (P_Z X_1, P_Z X_2) = (\hat{X}_1, \hat{X}_2)$, one easily finds that $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$, with

$$\hat{\beta}_2 = (\hat{X}_2' M_{\hat{X}_1} \hat{X}_2)^{-1} \hat{X}_2' M_{\hat{X}_1} y \tag{2.3}$$

and $\widehat{Var}(\hat{\beta}_2) = \hat{\sigma}_\varepsilon^2 (\hat{X}_2' M_{\hat{X}_1} \hat{X}_2)^{-1}$, where $M_A = I - P_A$. The W statistic for testing $\mathcal{H}_0 : \beta_2 = 0$ against alternative $\mathcal{H}_1 : \beta_2 \neq 0$ readily follows and is given by

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

Under $\mathcal{H}_0$ it is asymptotically $\chi^2(K_2)$ distributed, provided the maintained hypothesis does hold indeed. The maintained hypothesis (which is the union of $\mathcal{H}_0$ and $\mathcal{H}_1$ and all underlying assumptions) can here be characterized loosely as

$\mathcal{M}_1$*: (i) the actual data generating process (DGP) is nested in model (2.2) with* K regressors*; (ii)* X *and* Z *show sufficient regularity; (iii)* $n^{-1/2} Z'\varepsilon \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 \operatorname{plim} n^{-1} Z'Z)$.

Substituting (2.3) in (2.4) 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

$$W_{\beta_2} = y' P_{M_{\hat{X}_1} \hat{X}_2} y / \hat{\sigma}_\varepsilon^2 = y'(P_{\hat{X}} - P_{\hat{X}_1}) y / \hat{\sigma}_\varepsilon^2 = y'(M_{\hat{X}_1} - M_{\hat{X}}) y / \hat{\sigma}_\varepsilon^2. \tag{2.5}$$

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 σ_ε^2. If the latter is obtained from the restricted residuals $\tilde{\varepsilon} = y - X_1 \tilde{\beta}_1$, where

$$\tilde{\beta}_1 = (\hat{X}_1' \hat{X}_1)^{-1} \hat{X}_1' y \tag{2.6}$$

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

$$LM_{\beta_2} = y'(M_{\hat{X}_1} - M_{\hat{X}}) y / \tilde{\sigma}_\varepsilon^2, \tag{2.7}$$

which is asymptotically equivalent with W_{β_2}.

A test statistic with similarities to the LR principle is found as follows. Estimator $\hat{\beta}$ is obtained by minimizing with respect to β 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

$$Q(\beta; y, X, Z, \sigma_\varepsilon^2) = (y - X\beta)' Z (Z'Z)^{-1} Z' (y - X\beta) / \sigma_\varepsilon^2, \tag{2.8}$$

one can verify that it attains its minimum for $\hat{\beta}$ of (2.1), giving

$$\begin{aligned} Q(\hat{\beta}; y, X, Z, \sigma_\varepsilon^2) &= (y - X\hat{\beta})' P_Z (y - X\hat{\beta}) / \sigma_\varepsilon^2 = (y' P_Z y - 2\hat{\beta}'\hat{X} y + \hat{\beta}'\hat{X}'\hat{X}\hat{\beta}) / \sigma_\varepsilon^2 \\ &= (y' P_Z y - y' P_{\hat{X}} y) / \sigma_\varepsilon^2 = (y' M_{\hat{X}} y - y' M_Z y) / \sigma_\varepsilon^2, \end{aligned}$$

which has in its numerator the difference between the sum of the squares of the second-stage residuals and the reduced form residuals. In the model under $\beta_2 = 0$, upon using the criterion function $Q(\beta_1; y, X_1, Z, \sigma_\varepsilon^2)$, one finds $\tilde{\beta}_1$ of (2.6) with $Q(\tilde{\beta}_1; y, X_1, Z, \sigma_\varepsilon^2) = (y'M_{\hat{X}_1}y - y'M_Z y)/\sigma_\varepsilon^2$. The difference between these two minimized criterion functions is

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

Now substituting either $\hat{\sigma}_\varepsilon^2$ or $\tilde{\sigma}_\varepsilon^2$, we find that taking the difference between these two evaluated criterion functions yields either

$$CF_{\beta_2}(\hat{\sigma}_\varepsilon^2) = Q(\tilde{\beta}_1; y, X_1, Z, \hat{\sigma}_\varepsilon^2) - Q(\hat{\beta}; y, X, Z, \hat{\sigma}_\varepsilon^2) = W_{\beta_2} \tag{2.9}$$

or

$$CF_{\beta_2}(\tilde{\sigma}_\varepsilon^2) = LM_{\beta_2}. \tag{2.10}$$

These equivalences are algebraic, thus hold irrespective of the validity of $\mathcal{M}_1$.

3. The Sargan test and exclusion restrictions

A special situation occurs in case $L = K$. This specializes $\mathcal{M}_1$ to what we will indicate by $\mathcal{M}_1^{L=K}$. Under this maintained hypothesis the model is just identified, but overidentified under the K_2 coefficient restrictions $\beta_2 = 0$. When the model is just identified then, when minimizing $Q(\beta; y, X, Z, \sigma_\varepsilon^2)$, the $L = K$ orthogonality conditions $Z'(y - X\hat{\beta}) = 0$ will all be satisfied in the sample, giving $Q(\hat{\beta}; y, X, Z, \sigma_\varepsilon^2) = 0$. So, in this particular situation, (2.9) specializes into

$$W_{\beta_2}^{L=K} = CF_{\beta_2}^{L=K}(\hat{\sigma}_\varepsilon^2) = Q(\tilde{\beta}_1; y, X_1, Z, \hat{\sigma}_\varepsilon^2) = \tilde{\varepsilon}' P_Z \tilde{\varepsilon}/\hat{\sigma}_\varepsilon^2, \tag{3.1}$$

and (2.10) into the asymptotically equivalent statistic

$$LM_{\beta_2}^{L=K} = CF_{\beta_2}^{L=K}(\tilde{\sigma}_\varepsilon^2) = \tilde{\varepsilon}' P_Z \tilde{\varepsilon}/\tilde{\sigma}_\varepsilon^2. \tag{3.2}$$

The latter expression is the well-known Sargan test statistic which is used to check the validity of the instruments by testing the $L - K_1$ overidentifying restrictions in the model which has just the K_1 regressors X_1, and can be expressed as

$$y = X_1\beta_1^* + \varepsilon^* \text{ with } \varepsilon^* \sim (0, \sigma_{\varepsilon^*}^2 I). \tag{3.3}$$

From its algebraic equivalence with a CR test it is immediately obvious that testing for OR 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 $\mathcal{M}_1^{L=K}$ for the test of $L - K_1$ exclusion restrictions $\beta_2 = 0$ in model (2.2), whereas the Sargan test is used to verify validity of the orthogonality conditions $E(z_i\varepsilon_i^*) = 0$ in model (3.3). Although using exactly the same test statistic and critical values, we will explicate that the two tests have intrinsically different $\mathcal{H}_0$ and $\mathcal{H}_1$ and build on different maintained hypotheses. We are not aware of literature where these differences and their consequences have been spelled out in full detail. Below we aim to fill this gap.

Note that expression (3.2) is invariant regarding X_2. So, replacing X_2 by any alternative $n \times K_2$ matrix X_2^a would yield equivalent results for $CF_{\beta_2}^{L=K}(\tilde{\sigma}_\varepsilon^2)$, provided $P_Z(X_1, X_2^a)$ has full column rank $K = L$. Thus, (X_1, X_2^a) should have full column rank and the variables X_2^a should be sufficiently related to the instruments Z, but at the same time they may also be endogenous regarding ε. 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 CR test actually does not test the explanatory power of X_2^a as such, but just that of $M_{\hat{X}_1}P_ZX_2^a = (P_Z - P_{\hat{X}_1}P_Z)X_2^a = (P_Z - P_{\hat{X}_1})X_2^a$, which is its projection on the space spanned by Z, as far as this is orthogonal to P_ZX_1.

A viable choice for X_2^a would be the following. Let $F = (F_1, F_2)$ be a full rank $L \times L$ matrix, where F_1 is $L \times L_1$, F_2 is $L \times L_2$, with $L = L_1 + L_2$ and $L_1 = K_1$. Now consider the transformed instruments $Z^f = (Z_1^f, Z_2^f) = ZF = (ZF_1, ZF_2)$, and let F_1 be such that square matrix $Z_1^{f\prime}X_1$ has full rank. Obviously, F is nonunique. Note that if the transformed instruments Z_1^f are not just relevant for X_1, but were valid for the errors ε^* of model (3.3) as well, then the coefficients β_1^* would be just identified by them. When taking $X_2^a = M_{Z_1^f}Z_2^f$, then $P_Z(X_1, X_2^a)$ has full column rank as required, and statistic $CF_{\beta_2^{**}}^{L=K}(\tilde{\sigma}_\varepsilon^2)$ for testing $\mathcal{H}_0 : \beta_2^{**} = 0$ in model

$$y = X_1\beta_1^{**} + M_{Z_1^f}Z_2^f\beta_2^{**} + \varepsilon^{**} \tag{3.4}$$

is therefore equivalent to the Sargan test statistic $CF_{\beta_2}^{L=K}(\tilde{\sigma}_\varepsilon^2)$. Estimating β_1^{**} of (3.4) by IV using the instruments Z or Z^f (which is equivalent because $P_Z = P_{Z^f}$) yields

$$\begin{aligned}\hat{\beta}_1^{**} &= (X_1'P_{Z^f}M_{M_{Z_1^f}Z_2^f}P_{Z^f}X_1)^{-1}X_1'P_{Z^f}M_{M_{Z_1^f}Z_2^f}P_{Z^f}y \\ &= (X_1'P_{Z_1^f}X_1)^{-1}X_1'P_{Z_1^f}y = (Z_1^{f\prime}X_1)^{-1}Z_1^{f\prime}y = \hat{\beta}_1^*,\end{aligned} \tag{3.5}$$

because $P_{Z^f}M_{M_{Z_1^f}Z_2^f}P_{Z^f} = P_{Z^f} - P_{M_{Z_1^f}Z_2^f} = P_{Z_1^f}$. Hence, augmenting model (3.3) by K_2 regressors $M_{Z_1^f}Z_2^f$ and using instruments Z or Z^f yields coefficient estimates $\hat{\beta}_1^{**}$ for the regressors X_1 which are equivalent to the simple IV estimator $\hat{\beta}_1^*$ obtained in model (3.3) when just using the K_1 instruments Z_1^f, irrespective of the validity of any of the instruments.

Apart from the interpretation of $CF_{\beta_2}^{L=K}(\tilde{\sigma}_\varepsilon^2)$ as testing $K_2 = L - K_1$ exclusion restrictions $\beta_2 = 0$ in model (2.2) under maintained hypothesis $\mathcal{M}_1^{L=K}$, which presupposes the K orthogonality conditions $E(z_i\varepsilon_i) = 0$, yet another coherent interpretation of the numerically equivalent test statistic $CF_{\beta_2^{**}}^{L=K}(\tilde{\sigma}_\varepsilon^2)$ is the following. It tests whether the $L - K_1$ variables Z_2^f are additional valid external instruments for model (3.3) under the maintained hypothesis given by

$\mathcal{M}_2^{L=K}$*: (i) the DGP is nested in model (3.3) which has* $K_1 < K = L$ regressors*; (ii)* X_1 *and* Z *(and thus* Z^f*) show sufficient regularity; (iii)* $n^{-1/2}Z_1^{f\prime}\varepsilon^* \xrightarrow{d} \mathcal{N}(0, \sigma_{\varepsilon^*}^2 \operatorname{plim} n^{-1}Z_1^{f\prime}Z_1^f)$.

Note that for adopting $\mathcal{M}_2^{L=K}$ in practice it might be desirable to specify matrix F_1; this issue will be addressed below.

Under $\mathcal{M}_2^{L=K}$ the value of $E(z_{2i}^f\varepsilon_i^*) = \sigma_{z_2^f\varepsilon^*}$ is unspecified. Obviously, if next to Z_1^f the variables Z_2^f are uncorrelated with ε^* then $M_{Z_1^f}Z_2^f$ should not have explanatory power in model (3.4). Hence, when adopting $\mathcal{M}_2^{L=K}$, then under the null hypothesis $\beta^{**} = 0$ in model (3.4) all instruments Z are valid for model (3.3), implying $\sigma_{z_2^f\varepsilon^*} = 0$ so that $\tilde{\beta}_1$ is consistent, asymptotically normal and efficient for β_1. Below we will simply use the acronym AE (asymptotically efficient) to indicate for an estimator if, under the assumptions made, it is consistent with asymptotically normal distribution and smallest asymptotic variance (in a matrix sense). The alternative $\beta^{**} \neq 0$ implies $\sigma_{z_2^f\varepsilon^*} \neq 0$. So, in the context of model (3.4) test statistic $CF_{\beta_2^{**}}^{L=K}(\tilde{\sigma}_\varepsilon^2)$ tests the null hypothesis given by the $L - K_1$ orthogonality conditions $E(z_{2i}^f\varepsilon_i^*) = 0$ or overidentifying restrictions $\beta^{**} = 0$. When these conditions/restrictions are invalid then, given validity of $\mathcal{M}_2^{L=K}$, estimator $\hat{\beta}_1^* = (Z_1^{f\prime}X_1)^{-1}Z_1^{f\prime}y$ for β_1^* is AE in model (3.3). In the extended regression model (3.4) all variables in both Z_1^f and Z_2^f constitute valid instruments regarding ε^{**} under $\mathcal{M}_2^{L=K}$, because inclusion of the regressors $M_{Z_1^f}Z_2^f$ purges the disturbances from their correlation with Z_2^f. That this regression too produces the AE estimator $\hat{\beta}_1^*$ for the coefficients β_1, as is proved by (3.5), is because it constitutes a special case of the so-called "partialling out" result which will be reviewed at the end of the next section.

It is the case that truth of the null $\beta_2 = 0$ under $\mathcal{M}_1^{L=K}$ implies $E(z_i\varepsilon_i^*) = 0$, and so does truth of the null $E(z_{2i}^f\varepsilon_i^*) = 0$ under $\mathcal{M}_2^{L=K}$. Under both null hypotheses estimator $\tilde{\beta}_1$ is AE for β_1. However, imposing the respective null hypotheses implies either accepting coefficient restrictions or accepting extra orthogonality assumptions. Invalidity of $\beta_2 = 0$ under $\mathcal{M}_1^{L=K}$ renders $\hat{\beta}$ the AE estimator for β, but invalidity of $E(z_{2i}^*\varepsilon_i^*) = 0$ under $\mathcal{M}_2^{L=K}$ renders $\hat{\beta}_1^*$ AE for the coefficients of just the regressors X_1, since in this setup the DGP is supposed to be nested in model (3.3) which has just K_1 regressors.

So, on the basis of the single test statistic $LM_{\beta_2}^{L=K}(\tilde{\sigma}_\varepsilon^2) = CF_{\beta_2^{**}}^{L=K}(\tilde{\sigma}_\varepsilon^2)$, or the asymptotically equivalent $W_{\beta_2}^{L=K}(\hat{\sigma}_\varepsilon^2)$, one can discriminate between either (in)validity of $L - K_1$ exclusion restrictions in model (2.2) or (in)validity of $L - K_1$ external instruments in model (3.3). This just hinges upon which of the nonnested maintained hypotheses has actually been adopted: $\mathcal{M}_1^{L=K}$ or $\mathcal{M}_2^{L=K}$. In both contexts the tests concern a possible loss/gain in the degree of overidentification by $L - K_1$. Namely, either due to rejecting/accepting exclusion restrictions on the coefficients of potential explanatory variables X_2 in model (2.2), or on accepting/rejecting exclusion restrictions on the coefficients of the auxiliary regressors $M_{Z_1^f} Z_2^f$ in model (3.4). The latter implies accepting/rejecting validity of $L - K_1$ external instruments Z_2^f.

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 $E(z_i \varepsilon_i^*) = 0$ in model (3.3). 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, meaning: existence of nonlocal alternatives to the null that are not rejected with probability one asymptotically. The classic Sargan test on instrument validity is only consistent in case one is willing to adopt maintained hypothesis $\mathcal{M}_2^{L=K}$. It is uncomfortable but also unavoidable that this maintained hypothesis embraces the statistically untestable orthogonality conditions expressed by $E(z_{1i}^f \varepsilon_i^*) = 0$ $(i = 1, ..., n)$. These untestable conditions entail: It is possible to find K_1 linearly independent linear combinations of the K variables in matrix Z which establish valid instruments for model (3.3) and ensure just identification of β_1^*.

It is the researcher who may decide whether (s)he leaves these K_1 orthogonality conditions implicit, or makes them explicit by actually specifying the matrix F_1, for instance by choosing $F_1 = (I, O)'$ 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 $M_{Z_1} Z_2$ are valid too, which would imply $E(z_{i2} \varepsilon_i^*) = 0$ in model (3.3).

4. The incremental Sargan test and exclusion restrictions

The validity of a subset of the instruments Z_2^f for model (3.3) can be tested by a so-called incremental Sargan or IOR test. For that we will examine too in what way such a procedure has (dis)similarities with testing exclusion restrictions. However, for this analysis we will take as our starting point the usual model with K regressors and $L > K$ instruments and therefore have to introduce some further notation.

We consider a partition of the matrix of L 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 $L_a > 0$ instruments Z_a will be examined by testing. So the maintained hypothesis is now given by

$\mathcal{M}_2^{L>K}$: *(i) the DGP is nested in model* $y = X\beta + \varepsilon$, *which has* K *regressors; (ii)* X

and $Z = (Z_m, Z_a)$ *with* $L > L_m \geq K$ *show sufficient regularity; (iii)* $n^{-1/2}Z_m'\varepsilon \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 \operatorname{plim} n^{-1}Z_m'Z_m)$.

Just using the instruments Z_m yields estimator and residuals

$$\hat{\beta}_m = (X'P_{Z_m}X)^{-1}X'P_{Z_m}y, \ \hat{\varepsilon}_m = y - X\hat{\beta}_m. \tag{4.1}$$

From the foregoing section it follows that the Sargan statistic

$$S_m = n \cdot \hat{\varepsilon}_m' P_{Z_m}\hat{\varepsilon}_m / \hat{\varepsilon}_m'\hat{\varepsilon}_m \tag{4.2}$$

is asymptotically $\chi^2(L_m - K)$ distributed under $\mathcal{M}_2^{L>K}$ when $L_m > K$, whereas it is zero for $L_m = K$. Using the additional instruments as well yields $\hat{\beta}$ given in (2.1) and $\hat{\varepsilon} = y - X\hat{\beta}$ and enables to calculate

$$S_{ma} = n \cdot \hat{\varepsilon}'P_Z\hat{\varepsilon}/\hat{\varepsilon}'\hat{\varepsilon}. \tag{4.3}$$

Now the null hypothesis $E(z_{ia}\varepsilon_i) = 0$, which involves L_a orthogonality conditions for model (2.2) additional to the L_m orthogonality conditions already implied by $\mathcal{M}_2^{L>K}$, can be tested by the incremental Sargan statistic

$$IS_a = S_{ma} - S_m, \tag{4.4}$$

which under $\mathcal{M}_2^{L>K}$ has asymptotic null distribution $\chi^2(L_a)$. Note that for $L_m = K$ statistic IS_a corresponds to the Sargan test examined in the foregoing section, but now the role of X_1 is plaid by X, so K_1 is now K, whereas K_2 is given here by L_a and what earlier was called $\tilde{\varepsilon}$ is given by $\hat{\varepsilon}$ now.

Most published derivations of the null distribution of IS_a start off from a GMM (generalized method of moments) context. Our simple linear framework allows a very concise derivation, which goes as follows. Without loss of generality we assume $Z_m'Z_a = O$, which may have been achieved by replacing the original Z_a by $M_{Z_m}Z_a$. P_Z is not affected by this partial orthogonalization, but now $P_Z = P_{Z_m} + P_{Z_a}$. From $P_Z\hat{\varepsilon} = P_Z[I - X(X'P_ZX)^{-1}X'P_Z]\varepsilon = (P_Z - P_{P_ZX})\varepsilon$ we find $\hat{\varepsilon}'P_Z\hat{\varepsilon} = \varepsilon'P_Z\varepsilon - \varepsilon'P_{P_ZX}\varepsilon = \varepsilon'(P_{Z_m} + P_{Z_a} - P_{P_ZX})\varepsilon$ and $\hat{\varepsilon}_m'P_{Z_m}\hat{\varepsilon}_m = \varepsilon'(P_{Z_m} - P_{P_{Z_m}X})\varepsilon$, so $\hat{\varepsilon}'P_Z\hat{\varepsilon} - \hat{\varepsilon}_m'P_{Z_m}\hat{\varepsilon}_m = \varepsilon'(P_{Z_a} + P_{P_{Z_m}X} - P_{P_ZX})\varepsilon$. The matrix in this quadratic form is symmetric and idempotent. The latter follows from $P_{Z_a}P_{P_{Z_m}X} = O$ and $(P_{Z_a} + P_{P_{Z_m}X})P_{P_ZX} = (P_{Z_a} + P_{Z_m})X(X'P_ZZ)^{-1}X'P_Z = P_{P_ZX}$. Since $tr(P_{Z_a} + P_{P_{Z_m}X} - P_{P_ZX}) = L_a + K - K = L_a$ this means that under $\mathcal{M}_2^{L>K}$ and the null hypothesis $E(z_{ia}\varepsilon_i) = 0$ $(i = 1, ..., n)$, so when all instrument in Z are valid, $IS_a \xrightarrow{d} \varepsilon'(P_{Z_a} + P_{P_{Z_m}X} - P_{P_ZX})\varepsilon/\sigma_\varepsilon^2 \xrightarrow{d} \chi^2(L_a)$, because under the null both $\operatorname{plim} \hat{\varepsilon}_m'\hat{\varepsilon}_m/n = \sigma_\varepsilon^2$ and $\operatorname{plim} \hat{\varepsilon}'\hat{\varepsilon}/n = \sigma_\varepsilon^2$.

We shall now examine tests for omitted regressors (or exclusion restrictions) and seek cases which show correspondences with statistic IS_a. Consider the extended model with $K_a \leq L - K$ additional regressors X_a given by

$$y = X\beta + X_a\beta_a + \varepsilon_a, \text{ with } \varepsilon_a \sim (0, \sigma^2_{\varepsilon_a} I), \tag{4.5}$$

and the maintained hypothesis defined by

$\mathcal{M}_1^{L>K}$: *(i) the DGP is nested in model* $y = X\beta + X_a\beta_a + \varepsilon_a$, *which has* $K + K_a \leq L$ *regressors; (ii)* (X, X_a) *and* Z *show sufficient regularity; (iii)* $n^{-1/2}Z'\varepsilon_a \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\varepsilon_a} \operatorname{plim} n^{-1}Z'Z)$.

Let us indicate the IV results when estimating model (4.5) using instruments Z as $y = X\hat{\beta}_* + X_a\hat{\beta}_a + \hat{\varepsilon}_a$. Then the LM-like test statistic for $\mathcal{H}_0 : \beta_a = 0$ is given by

$$CF_{\beta_a}(\hat{\sigma}^2_\varepsilon) = (\hat{\varepsilon}'P_Z\hat{\varepsilon} - \hat{\varepsilon}_a'P_Z\hat{\varepsilon}_a)/\hat{\sigma}^2_\varepsilon, \tag{4.6}$$

which under $\mathcal{M}_1^{L>K}$ has asymptotic null distribution $\chi^2(K_a)$.

Still assuming without loss of generality that $Z_m'Z_a = O$, we shall first examine the specific case $X_a = Z_a$, for which expression (4.6) can be specialized by using the following results. Substituting $X_a = Z_a$ we obtain

$$\hat{\beta}_* = (X'P_Z M_{Z_a} P_Z X)^{-1}X'P_Z M_{Z_a} y = (X'P_{Z_m}X)^{-1}X'P_{Z_m}y = \hat{\beta}_m, \tag{4.7}$$

because $M_{Z_a}P_Z = P_Z - P_{Z_a}P_Z = P_Z - P_{Z_a} = P_{Z_m}$. And, since in IV estimation the residuals are orthogonal to the second-stage regressors, we have $Z_a'\hat{\varepsilon}_a = 0$. This yields $P_Z\hat{\varepsilon}_a = (P_{Z_m} + P_{Z_a})\hat{\varepsilon}_a = P_{Z_m}(y - X\hat{\beta}_* - Z_a\hat{\beta}_a) = P_{Z_m}(y - X\hat{\beta}_*) = P_{Z_m}(y - X\hat{\beta}_m) = P_{Z_m}\hat{\varepsilon}_m$, where we used (4.7) for the last equality. From this we find

$$\begin{aligned} CF_{\beta_a}^{X_a=Z_a}(\hat{\sigma}^2_\varepsilon) &= (\hat{\varepsilon}'P_Z\hat{\varepsilon} - \hat{\varepsilon}_a'P_Z\hat{\varepsilon}_a)/\hat{\sigma}^2_\varepsilon = (\hat{\varepsilon}'P_Z\hat{\varepsilon} - \hat{\varepsilon}_m'P_{Z_m}\hat{\varepsilon}_m)/\hat{\sigma}^2_\varepsilon \\ &= S_{ma} - S_m\frac{\hat{\varepsilon}_m'\hat{\varepsilon}_m}{\hat{\varepsilon}'\hat{\varepsilon}}. \end{aligned} \tag{4.8}$$

Under $\mathcal{M}_1^{L>K}$ and $\mathcal{H}_0 : \beta_a = 0$ both $\hat{\varepsilon}_m'\hat{\varepsilon}_m/n$ and $\hat{\varepsilon}'\hat{\varepsilon}/n$ converge to σ^2_ε, thus the ratio $\hat{\varepsilon}_m'\hat{\varepsilon}_m/\hat{\varepsilon}'\hat{\varepsilon}$ equals 1 asymptotically. Therefore, under their respective null hypotheses, the incremental Sargan test statistic $IS_a = S_{ma} - S_m$ is asymptotically equivalent with the exclusion restrictions test statistic $CF_{\beta_a^*}^{X_a=Z_a}(\hat{\sigma}^2_\varepsilon)$.

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

$$X_a^* = Z_aC_1 + XC_2 + M_ZC_3 + \varepsilon C_4, \tag{4.9}$$

where the finite matrices C_1, C_2, C_3, C_4 are $L_a \times K_a$, $K \times K_a$, $n \times K_a$ and $1 \times K_a$ respectively. Now writing $\hat{X}_a^* = P_ZX_a^*$ and $\hat{X} = P_ZX$ it follows that $\hat{X}_a^* = Z_aC_1 + \hat{X}C_2 + P_Z\varepsilon C_4$. Under the null hypothesis $\beta_a = 0$, we have $\operatorname{plim} n^{-1}Z'\varepsilon = 0$, hence

$P_Z\varepsilon C_4 \rightarrow O$, thus $M_{\hat{X}}\hat{X}_a^* \rightarrow M_{\hat{X}}Z_aC_1$. The regularity assumption of $\mathcal{M}_1^{L>K}$ requires this to have full column rank, hence C_1 should have rank $K_a \leq L_a$. If $K_a = L_a$ and C_1 has full rank we obtain $P_{M_{\hat{X}}\hat{X}_a^*} \rightarrow P_{M_{\hat{X}}Z_a}$. Thus, tests for the omission of regressors X_a which belong to group (4.9) with C_1 of full rank $K_a = L_a$ will yield test statistics which under the null are all asymptotically equivalent with statistic IS_a.

Estimator $\hat{\beta}$ is AE both under $\mathcal{M}_1^{L>K}$ with valid exclusion restrictions $\beta_a = 0$ and under $\mathcal{M}_2^{L>K}$ with valid additional instruments Z_a. However, under $\mathcal{M}_2^{L>K}$ while instruments Z_a are invalid $\hat{\beta}_m$ is AE, whereas under $\mathcal{M}_1^{L>K}$ with $\beta_a \neq 0$ the estimators $\hat{\beta}_*$ and $\hat{\beta}_a$ are AE in model (4.5), yielding an estimator for β_* asymptotically equivalent to $\hat{\beta}_m$ only if X_a belongs to group (4.9) with $K_a = L_a$, C_1 of full rank and $C_2 = O$.

The algebraic estimator equivalence for $X_a = Z_a$ given in (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 (in our notation) $\hat{\beta}_*$ and $\hat{\beta}_m$ "hardly differ". Moreover, in the simulation design used 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 Neyman-Pearson statistical test methodology. In practice, however, a more flexible though opportunistic approach may often seem more appealing. The two juxtaposed maintained hypotheses $\mathcal{M}_1^{L=K}$ and $\mathcal{M}_2^{L=K}$ in Section 3, and their generalizations $\mathcal{M}_1^{L>K}$ and $\mathcal{M}_2^{L>K}$ 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[1] when obtaining a significant OR test may either conclude: (a) some instruments are invalid, (b) the model specification is deficient (so its implied errors may correlate even with instruments which would be valid for an adequate model specification), or (c) both model and instruments are unsound. We have shown that this shilly-shally is a consequence of not building on a firm maintained hypothesis. Under $\mathcal{M}_2^{L>K}$ an insignificant OR statistic is either due to validity of the tested additional moment conditions, or to lack of power of the test under (a), but not to (b) or (c), whereas a significant OR statistic is either due to a Type I error or to invalid instruments. Prac-

[1] See, for instance, Hayashi (2000, p.218) and Cameron and Trivedi (2005, p.277).

titioners who interpret an insignificant OR statistic as approval of the chosen model specification and the employed instruments should always realize that this presupposes validity of the untested hypothesis that this instrument set contains a subset which already just-identifies the coefficients of this chosen model specification.

The presented results on instrument validity and coefficient restriction tests can rather straight-forwardly be generalized to linear models with nonspherical disturbances estimated by generalized method of moments, because GMM conforms to IV/2SLS after proper transformations. When $\varepsilon \sim (0, \sigma_\varepsilon^2 \Omega)$ this transformation requires (as in GLS) to premultiply the model by an $n \times n$ matrix Ψ, where $\Psi'\Psi = \Omega^{-1}$, so that $\varepsilon^* = \Psi\varepsilon \sim (0, \sigma_\varepsilon^2 I)$, but employ now as instruments $Z^\dagger = (\Psi')^{-1}Z$, see Kiviet and Feng (2016). Hence, all above results can easily be reformulated for the linear GMM context, where (incremental) OR tests are usually addressed as Sargan-Hansen tests and indicated as (differences) in J tests.[2] In case of unknown heteroskedasticity (hence Ω is diagonal but not available), rather than robustifying inefficient IV, its findings $\hat{\beta}$ and $\hat{\varepsilon} = y - X\hat{\beta}$ can be employed in a second step to obtain asymptotically efficient feasible GMM inference. This can be represented[3] as follows. Let $y_i^* = y_i/\hat{\varepsilon}_i$, $X^* = (x_1^*, .., x_n^*)'$ with $x_i^* = x_i/\hat{\varepsilon}_i$ and $Z^\dagger = (z_1^\dagger, ..., z_n^\dagger)$ with $z_i^\dagger = z_i \times \hat{\varepsilon}_i$. Next regressing y^* on X^* using instruments $Z^\dagger$ yields $\hat{\beta}^\dagger = (X^{*\prime} P_{Z^\dagger} X^*)^{-1} X^{*\prime} P_{Z^\dagger} y^* = [X'Z(Z^{\dagger\prime}Z^\dagger)^{-1}Z'X]^{-1}X'Z(Z^{\dagger\prime}Z^\dagger)^{-1}Z'y$, with $Z^{\dagger\prime}Z^\dagger = \Sigma_{i=1}^n \hat{\varepsilon}_i^2 z_i z_i'$. Its variance could be estimated consistently by $\widehat{Var}(\hat{\beta}^\dagger) = [X'Z(Z^{\dagger\prime}Z^\dagger)^{-1}Z'X]^{-1}$, and writing $\hat{\varepsilon}^\dagger = y^* - X^*\hat{\beta}^\dagger$ the Sargan-Hansen statistic is $S^\dagger = \hat{\varepsilon}^{\dagger\prime} P_{Z^\dagger} \hat{\varepsilon}^\dagger = (y - X\hat{\beta}^\dagger)'Z(Z^{\dagger\prime}Z^\dagger)^{-1}Z'(y - X\hat{\beta}^\dagger)$, which is asymptotically equivalent to the GMM maximized criterion function. Hence, also for feasible GMM all earlier findings on CR and IOR tests have their (asymptotically equivalent) counterparts. Note, however, that a robustified variance estimator for $\hat{\beta}$ would be $(\hat{X}'\hat{X})^{-1}\Sigma_{i=1}^n \hat{\varepsilon}_i^2 \hat{x}_i \hat{x}_i' (\hat{X}'\hat{X})^{-1}$ with $P_Z X = \hat{X} = (\hat{x}_1, ..., \hat{x}_n)'$. Employing this in a robustified version of the CR test W_{β_2} of (2.4) does no longer seem to lead to a statistic which can be expressed equivalently as the difference between maximized robustified criterion functions, as in (2.9). Note that a robustified Sargan statistic for IV is given by $\hat{\varepsilon}'(\Sigma_{i=1}^n \hat{\varepsilon}_i^2 z_i z_i')^{-1}\hat{\varepsilon} = (y - X\hat{\beta})'Z(Z^{\dagger\prime}Z^\dagger)^{-1}Z'(y - X\hat{\beta})$. Due to the consistency of $\hat{\beta}$ this is asymptotically equivalent[4] to the 2-step statistic $S^\dagger$, and thus a robustified CF test has correspondences with a CR test based on 2-step feasible GMM estimator $\hat{\beta}^\dagger$ and no longer with one based on the inefficient $\hat{\beta}$.

Also in the extended contexts indicated in the preceding paragraph the practical relevance of the here presented findings is the following. An (incremental) Sargan-Hansen

[2] When embedded into a properly extended framework generalizations for nonlinear models can be obtained as well; see, for instance, Newey (1985).

[3] See Newey and West (1987, p.786).

[4] Baum et al (2003, p.18) claim that S^* and $S^\dagger$ are even numerically equivalent, which does not seem right.

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 regressor variables which may stem from a particular fairly wide group. However, in essence these two test procedures build on different nonnested maintained hypotheses which enables them to produce discriminative inferences.

References

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

Cameron, A.C., Trivedi, P.K., 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

Guggenberger, P., 2012. A note on the (in)consistency of the test of overidentifying restrictions and the concepts of true and pseudo-true parameters. *Economics Letters* 117, 901-904.

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

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

Hendry, D.F., 2011. On adding over-identifying instrumental variables to simultaneous equations. *Economics Letters* 111, 68-70.

Kiviet, J.F., Feng, Q., 2016. Efficiency gains by modifying GMM estimation in linear models under heteroskedasticity. UvA-Econometrics working paper 2014/06.

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

Newey, W.K., West, K.D., 1987. Hypothesis testing with efficient method of moments estimation. *International Economic Review* 28, 777-787.

Parente, P.M.D.C, Santos Silva, J.M.C., 2012. A cautionary note on tests of overidentifying restrictions. *Economics Letters* 115, 314-317.

Sargan, J.D., 1958. Estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-514.