Testing Over- and Underidentification in Linear Models, with Applications to Dynamic Panel Data and Asset-Pricing Models

Frank Windmeijer

Discussion Paper 18 / 696

19 March 2018

Department of Economics
University of Bristol
Priory Road Complex
Bristol BS8 1TU
United Kingdom

Testing Over- and Underidentification in Linear Models, with Applications to Dynamic Panel Data and Asset-Pricing Models*

Frank Windmeijer[†]
Department of Economics and IEU
University of Bristol, UK
and Cemmap, UCL/IFS, London, UK

March 2018

Abstract

This paper develops the links between overidentification tests, underidentification tests, score tests and the Cragg-Donald (1993, 1997) and Kleibergen-Paap (2006) rank tests in linear instrumental variables (IV) models. This general framework shows that standard underidentification tests are (robust) score tests for overidentification in an auxiliary linear model, $x_1 = X_2\delta + \varepsilon_1$, where $X = [x_1\ X_2]$ are the endogenous explanatory variables in the original model, estimated by IV estimation methods using the same instruments as for the original model. This simple structure makes it possible to establish valid robust underidentification tests for linear IV models where these have not been proposed or used before, like clustered dynamic panel data models estimated by GMM. The framework also applies to general tests of rank, including the I test of Arellano, Hansen and Sentana (2012), and, outside the IV setting, for tests of rank of parameter matrices estimated by OLS. Invariant rank tests are based on LIML or continuously updated GMM estimators of the first-stage parameters. This insight leads to the proposal of a new two-step invariant asymptotically efficient GMM estimator, and a new iterated GMM estimator that converges to the continuously updated GMM estimator.

JEL Classification: C12, C13, C23, C26
Keywords: Overidentification, Underidentification, Rank tests, Dynamic Panel Data Models, Asset Pricing Models

---

*This research was partly funded by the Medical Research Council (MC_UU_12013/9). I would like to thank Peter Boswijk, Jason Dovonon, Jean-Marie Dufour, Bo Honoré, Pascal Lavergne, Mark Schaffer, Enrique Sentana, Jon Temple and seminar and conference participants at Amsterdam, Bristol, Canterbury, Durham, Hong Kong, Lisbon, Melbourne, Montreal, Oxford, Toulouse and Perth for helpful comments.

†E-mail address: f.windmeijer@bristol.ac.uk.

1 Introduction

It is common practice when reporting estimation results of standard linear instrumental variables (IV) models to include the first-stage F, Cragg-Donald (Cragg and Donald, 1993, 1997) and/or Kleibergen-Paap (Kleibergen and Paap, 2006) test statistics. These are underidentification tests, testing the null hypothesis that the instruments have insufficient explanatory power to predict the endogenous variable(s) in the model for identification of the parameters. In the linear projection model for the endogenous explanatory variables X on the instruments Z, $X = Z\Pi + V$, they are tests on the rank of Π, with the standard tests testing the null, $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$, where k_x is the number of explanatory variables. Partition $X = \left[\begin{array}{cc} x_1 & X_2 \end{array} \right]$, then we show that these underidentification tests are tests for overidentification in the auxiliary model $x_1 = X_2\delta + \varepsilon_1$, estimated by IV methods using Z as instruments. The non-robust Cragg-Donald statistic is then equal to the Sargan (1958) or Basmann (1960) tests for overidentifying restrictions after estimating the parameters in the auxiliary model by limited information maximum likelihood (LIML). A version of the robust Cragg-Donald statistic is the Hansen J-test (Hansen, 1983), based on the continuously updated generalised method of moments (CU-GMM) estimator. Robustness here is with respect to the variance of the limiting distribution of $Z'\varepsilon_1/\sqrt{n}$, and robust to heteroskedasticity, time series correlation and/or clustering. These LIML and CU-GMM estimators and tests are invariant to the choice of normalisation (or choice x_j as the dependent variable) in the auxiliary regression.

We further show that the robust Kleibergen-Paap test is a LIML based invariant robust score test. In order to develop these arguments, we first discuss general testing for overidentifying restrictions for the standard linear model of interest $y = X\beta + u$, estimated by IV methods using instruments Z. Following Davidson and MacKinnon (1993), we show that (robust) tests for overidentifying restrictions are (robust) score tests for the null $H_0 : \gamma = 0$ in the specification $y = X\beta + Z_2\gamma + u$, where Z_2 is any $k_z - k_x$ subset of instruments. As this is a just identified model, it follows that the local asymptotic power of the tests is not affected by the choice of IV estimator in the restricted model. We further show that the standard robust two-step GMM estimator based Hansen J-test is a robust score test which is a function of the data and the one-step GMM estimator only. The one-step GMM estimator enters the Hansen J-test through the consistent variance estimator, and hence the local asymptotic power of the test is not affected by the choice of one-step estimator.

The robust Kleibergen-Paap and Cragg-Donald statistics as tests for overidentification are robust score tests for $H_0 : \gamma = 0$. The standard Kleibergen-Paap robust score test is based on the LIML estimators for β and Π in the restricted model, whereas the Cragg-Donald robust score test is based on the CU-GMM estimators for β and Π. They achieve invariance by incorporating the LIML or CU-GMM estimator for Π to form the optimal combination of instruments. This differs from the standard two-step GMM framework which uses the OLS estimator for Π in both one-step and two-step estimators for β. We use this observation to propose a two-step invariant asymptotically efficient GMM estimator that is based on the LIML estimator for β as the one-step estimator and uses the LIML estimator for Π to construct the optimal instruments for the second step. Alternatively, one can update the estimator for Π from the first-order conditions of the CU-GMM estimator, which leads to a different invariant two-step estimator and an iterated GMM estimator that converges to the CU-GMM estimator, also when starting from any non-invariant one-step estimator.

Consider the linear projection $\left[\begin{array}{cc} y & X \end{array} \right] = Z \left[\begin{array}{cc} \pi_y & \Pi \end{array} \right] + \left[\begin{array}{cc} v_y & V \end{array} \right] = Z\Pi^* + V^*$, then the Kleibergen-Paap and Cragg-Donald statistics for overidentifying restrictions are invariant rank tests for $H_0 : \mathrm{r}(\Pi^*) = k_x$ against $H_1 : \mathrm{r}(\Pi^*) = k_x + 1$. We therefore establish here the link between the test for overidentifying restrictions, score tests and rank tests. This carries over directly to the tests for underidentification, which are (robust) tests for $H_0 : \gamma = 0$ in the specification $x_1 = X_2\delta + Z_2\gamma + \varepsilon_1$, using generic notation for Z_2 and γ. Here, Z_2 is any $k_z - k_x + 1$ subset of instruments.

Instead of a single invariant underidentification test, Sanderson and Windmeijer (2016) considered per endogenous explanatory variable non-invariant tests. These are tests for overidentification in the k_x specifications $x_j = X_{-j}\delta_j + \varepsilon_j$, where X_{-j} is X without x_j. In the homoskedastic case, these are 2SLS based non-robust Sargan or Basmann tests. Robust tests are then easily obtained as two-step GMM Hansen J-tests.

For the homoskedastic case, the non-robust Cragg-Donald and Sanderson-Windmeijer tests can be used for testing for weak instruments. This is the framework developed by Staiger and Stock (1997) and Stock and Yogo (2005) and considers the situation where the test for underidentification rejects the null, but the information content of the instruments is such that the IV estimator is biased and the Wald test on the structural parameters is size distorted. This leads to larger critical values for the tests with the null hypotheses now being in terms of the maximal relative bias of the IV estimator, relative to that of the OLS estimator and/or the size distortion of the Wald test. However, the Stock and Yogo (2005) weak-instrument test results do not apply when using robust test statistics for when the errors are conditionally heteroskedastic, correlated over time and/or clustered, see Bun and De Haan (2010), Andrews (2017) and Kim (2017).

Linear dynamic panel data models for panels with a large cross-sectional dimension n and short time series dimension T are a leading example of linear IV models with clustered and potentially heteroskedastic data. The commonly used Arellano and Bond (1991) and Blundell and Bond (1998) estimation procedures are robust one-step, or efficient two-step GMM estimators, where lagged levels are instruments for first-differences of economic series, or lagged first-differences are instruments for levels. Whilst the problem of weak instruments for these models have been well documented and are mainly due to the persistence over time of many economic series, see e.g. Blundell and Bond (1998), Bun and Windmeijer (2010) and Hayakawa and Qi (2017), estimation results rarely include test results for underidentification. Bazzi and Clemens (2013) considered the Kleibergen-Paap test for this setting, but only in a per period cross-sectional setting, following the analysis of Bun and Windmeijer (2010). This is not a valid approach for testing for underidentification as it does not take clustering into account.

From the exposition above, robust testing for underidentification in dynamic panel data models is quite straightforward. For example for the first-differenced Arellano-Bond procedure, the auxiliary model is $\Delta x_1 = (\Delta X_2)\,\delta + \varepsilon_1$, and the robust Cragg-Donald statistic is the J-test based on the cluster robust CU-GMM estimator for δ. The non-robust Cragg-Donald statistic needs to be obtained here with iterative methods due to the clustered nature of the errors. The robust Kleibergen-Paap test does not need iterative methods as it is the cluster robust score test based on the pooled LIML estimator. The robust Sanderson-Windmeijer tests are simply the per endogenous explanatory variables Hansen two-step J-tests. The latter are particularly easy to compute as the estimation procedure of the auxiliary model is the same as that of the original model.

The relationships between rank tests, score tests and overidentification tests readily extend to testing for general rank, which establishes a direct link with the underidentification test for linear models as proposed by Arellano, Hansen and Sentana (2012). The results are also not specific to IV models, but apply to general settings of parameter matrices that are estimated by OLS, as considered by for example Al-Sadoon (2017). Even within this OLS setting, the Cragg-Donald and Kleibergen-Paap rank tests are LIML/CU-GMM based overidentification tests. For example, in the model $Y = XB + U$, with X now exogenous variables and $k_x \geq k_y$, for testing $H_0 : \mathrm{r}(B) = k_y - 1$ against $H_1 : \mathrm{r}(B) = k_y$, the auxiliary model is $y_1 = Y_2\delta + \varepsilon_1$ estimated by IV methods using X as the instruments. We show that when $k_x \leq k_y$, and for testing $H_0 : \mathrm{r}(B') = k_x - 1$ against $H_1 : \mathrm{r}(B') = k_x$, the auxiliary model is $x_1 = X_2\delta + \varepsilon_1$ with Y as the instruments.

The latter representation fits the setting of linear asset pricing factor models when the number of asset returns is larger than the number of factors, as considered recently by Gospodinov, Kan and Robotti (2017). They consider the robust Cragg-Donald rank test/CU-GMM J-test for overidentifying restrictions. They derive the limiting distribution of the test for overidentifying restrictions in underidentified models under homoskedasticity. We show that their result applies directly to the LIML based Sargan test. It follows that the test has no power to detect invalid overidentifying restrictions in underidentified models. This is therefore clearly another important reason to report underidentification tests. Also, the underidentification test is not adversely effected if the matrix is of lower rank than the null, as it will be undersized and hence does not lead to erroneous conclusions. We repeat the Monte Carlo analysis of Gospodinov, Kan and Robotti (2017) and find that the underidentification tests perform well, with the Sanderson-Windmeijer tests able to identify the spurious factors that cause underidentification.

The structure of the paper is as follows, Section 2 introduces the linear IV model and assumptions. Section 3 derives the results for the tests for overidentifying restrictions, establishes the links with score tests and rank tests, and introduces the invariant two-step GMM estimator and iterated CU-GMM estimator. Section 4 discusses the tests for underidentification and Section 5 develops these for use in dynamic panel data models. Section 6 extends the analysis to testing for general rank and links to the testing procedure of Arellano, Hansen and Sentana (2012). Section 7 generalises the results to rank tests outside the linear IV setting. Section 8 considers the limiting distribution of the Sargan test in underidentified models and Section 9 repeats the Monte Carlo analysis of Gospodinov, Kan and Robotti (2017), but includes testing for underidentification. Section 10 concludes.

2 Model and Assumptions

We consider the linear model

$$y = X\beta + u, \tag{1}$$

where y and u are the n-vectors (y_i) and (u_i), and X is the $n \times k_x$ matrix $[x_i']$, for the sample $i = 1, .., n$. The explanatory variables are endogenous, and Z is an $n \times k_z$ matrix $[z_i']$ of instrumental variables, with $k_z > k_x$. Note that exogenous explanatory variables have been partialled out. The first-stage, linear projection for X is given by

$$X = Z\Pi + V, \tag{2}$$

where Π is a $k_z \times k_x$ matrix, and V the $n \times k_x$ matrix $[v_i']$.

We make the following standard assumptions, see e.g. Stock and Yogo (2005),

Assumption 1 $E\left(z_i z_i'\right) = Q_{zz}$. *$Q_{zz}$ is nonsingular.*

Assumption 2 $E\left(z_i u_i\right) = 0$.

Assumption 3 $E\left(z_i x_i'\right) = Q_{zx}$ *has rank* k_x.

Assumption 4 $E\left[\begin{pmatrix} u_i \\ v_i \end{pmatrix} \begin{pmatrix} u_i & v_i' \end{pmatrix}\right] = \Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv}' \\ \sigma_{uv} & \Sigma_v \end{bmatrix}$.

Assumption 5 $\text{plim}\left(\frac{1}{n}Z'Z\right) = Q_{zz}$; $\text{plim}\left(\frac{1}{n}Z'X\right) = Q_{zx}$;

$$\text{plim}\left[\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} u_i \\ v_i \end{pmatrix} \begin{pmatrix} u_i & v_i' \end{pmatrix}\right] = \Sigma;$$

$$\frac{1}{\sqrt{n}}Z'u \to N\left(0, \Omega_{zu}\right); \ \frac{1}{\sqrt{n}}\text{vec}\left(Z'V\right) \xrightarrow{d} N\left(0, \Omega_{zv}\right).$$

These assumptions can hold for cross-sectional or time series data. As in Arellano, Hansen and Sentana (2012), time series data are stationary and ergodic. If the observations are independent/not correlated and conditionally homoskedastic, $E\left(u_i^2|z_i\right) = \sigma_u^2$, then $\Omega_{zu} = \sigma_u^2 Q_{zz}$. Throughout, we refer to this case simply as the homoskedastic case. We defer discussion of a clustered design to Section 5 on dynamic panel data models.

Moment restriction $E(z_i u_i) = 0$, Assumption 2, is the exclusion restriction that is generally tested by a test of overidentifying restrictions, like Hansen's J-test. Assumption 3, $E(z_i x_i')$ has full column rank k_x, is the relevance condition, which is commonly tested by a rank test like the Cragg-Donald (1993) statistic, which tests $H_0 : \mathrm{r}(E(z_i x_i')) = k_x - 1$ against $H_1 : \mathrm{r}(E(z_i x_i')) = k_x$. If $\mathrm{r}(E(z_i x_i')) = k_x - 1$, then there is a δ^*, such that $E(z_i x_i' \delta^*) = 0$. The similarity of testing procedures for over- or underidentification is then easily seen as, defining $w_i = \begin{pmatrix} y_i & x_i' \end{pmatrix}'$ and $\psi = \begin{pmatrix} 1 & -\beta' \end{pmatrix}'$, $E(z_i u_i) = E(z_i w_i' \psi) = 0$. Therefore, partitioning $X = \begin{bmatrix} x_1 & X_2 \end{bmatrix}$, a test for underidentification is an overidentification test for $H_0 : E(z_i \varepsilon_{1i}) = 0$ in the model $x_1 = X_2 \delta + \varepsilon_1$, with the test invariant to which explanatory variable is chosen as the dependent variable when it is based on an invariant estimator like LIML. Equivalently, a test for overidentification is a test for $H_0 : \mathrm{r}(E(z_i w_i')) = k_x$ against $H_1 : \mathrm{r}(E(z_i w_i')) = k_x + 1$.

Throughout the paper, we use the following notation for projection matrices. For a full column rank $n \times k_A$ matrix A, the projection matrix P_A is defined as $P_A = A(A'A)^{-1} A'$ and $M_A = I_n - P_A$.

3 Overidentification Tests

Under Assumptions 1-5, standard IV estimators for β, like 2SLS, LIML, GMM and CU-GMM are consistent and asymptotically normally distributed. The test for overidentifying restrictions is a test for $H_0 : E(z_i u_i) = 0$, and is a score test for the hypothesis $H_0 : \gamma = 0$ in the model

$$y = X\beta + Z_2\gamma + u, \tag{3}$$

where Z_2 is any $k_z - k_x$ subset of instruments, see e.g. Davidson and MacKinnon (1993, p. 235). The score test is invariant to the choice of instruments included in Z_2, unlike the Wald test. However, as (3) is a just identified model, different IV estimators all produce the same robust Wald test, the one based on the just-identified IV/2SLS estimator. The robust score and Wald tests have the same local asymptotic power, i.e. for alternatives $\gamma = c/\sqrt{n}$, see e.g. Wooldridge (2010, p 417). As asymptotically valid robust tests for overidentifying restrictions are robust score tests, it therefore follows that all tests have the same local asymptotic power, independent of the estimator the overidentification test is based on.

We will first discuss and derive some results for the standard two-step GMM Hansen J-test, in particular showing that it is equal to the robust score test based on a one-step estimator.

The one-step GMM estimator for β in model (1) is given by

$$\widehat{\beta}_1 = \left(X'ZW_n^{-1}Z'X\right)^{-1} X'ZW_n^{-1}Z'y,$$

where W_n is such that $n^{-1}W_n \xrightarrow{p} W$, a finite and positive definite matrix. The 2SLS estimator is a one-step GMM estimator with $W_n = Z'Z$. The one-step residual is given by $\widehat{u}_1 = y - X\widehat{\beta}_1$. The two-step GMM estimator is asymptotically efficient under general forms of heteroskedasticity and serial correlation, and is given by

$$\widehat{\beta}_2 = \left(X'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'X\right)^{-1} X'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'y,$$

where $Z'H_{\widehat{u}_1}Z$ is an estimator for Ω_{zu} such that $n^{-1}Z'H_{\widehat{u}_1}Z \xrightarrow{p} \Omega_{zu}$. For example, a Newey-West estimator robust to conditional heteroskedasticity and autocorrelation is given by

$$Z'H_{\widehat{u}_1}Z = \Gamma_{\widehat{u}_1}(0) + \sum_{l=1}^{p}\left(1 - \frac{l}{p+1}\right)\left(\Gamma_{\widehat{u}_1}(l) + \Gamma_{\widehat{u}_1}(l)'\right),$$

where $\Gamma_{\widehat{u}_1}(0) = \sum_{i=1}^{n} \widehat{u}_{1i}^2 z_i z_i'$, and $\Gamma_{\widehat{u}_1}(l) = \sum_{i=l+1}^{n} \widehat{u}_{1i}\widehat{u}_{1,i-l} z_i z_{i-l}'$.

Let $\widehat{u}_2 = y - X\widehat{\beta}_2$, then the Hansen J-test is given by

$$J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) = \widehat{u}_2'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'\widehat{u}_2, \tag{4}$$

and $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z-k_x}$ under Assumptions 1-5.

Let $\widehat{\Pi} = \left(Z'Z\right)^{-1}Z'X$ be the OLS estimator of Π, and $\widehat{X} = Z\widehat{\Pi}$. The following Proposition shows the equivalence of the robust score test and the J-test.

Proposition 1 *The robust score test for $H_0 : \gamma = 0$ in model (3) based on a one-step GMM estimator $\widehat{\beta}_1$ is given by*

$$\begin{aligned} S_r\left(\widehat{\beta}_1\right) &= \widehat{u}_1' M_{\widehat{X}} Z_2 \left(Z_2' M_{\widehat{X}} H_{\widehat{u}_1} M_{\widehat{X}} Z_2\right)^{-1} Z_2' M_{\widehat{X}} \widehat{u}_1 \\ &= y' M_{\widehat{X}} Z_2 \left(Z_2' M_{\widehat{X}} H_{\widehat{u}_1} M_{\widehat{X}} Z_2\right)^{-1} Z_2' M_{\widehat{X}} y, \end{aligned}$$

with $S_r\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z-k_x}$ under Assumptions 1-5.
Let the Hansen J-test $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right)$ be as defined in (4). Then

$$J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) = S_r\left(\widehat{\beta}_1\right).$$

Proof. See Appendix ∎

This result clearly shows that the two-step estimator is irrelevant in the calculation of the score test, and that the choice of one-step estimator does not affect the local asymptotic power of the test, as the one-step estimator enters the score test only through the variance estimator.

The non-robust version of the score test is given by

$$S\left(\widehat{\beta}_1\right) = \frac{\widehat{u}_1' M_{\widehat{X}} Z_2 \left(Z_2' M_{\widehat{X}} Z_2\right)^{-1} Z_2' M_{\widehat{X}} \widehat{u}_1}{\widehat{u}_1' \widehat{u}_1 / n},$$

with $S\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under Assumptions 1-5 in the homoskedastic case. From standard score test theory, for the efficient 2SLS estimator the score test becomes

$$S\left(\widehat{\beta}_{2sls}\right) = \frac{\widehat{u}_{2sls}' P_Z \widehat{u}_{2sls}}{\widehat{u}_{2sls}' \widehat{u}_{2sls} / n},$$

which also follows directly from the fact that $\widehat{X}' \widehat{u}_{2sls} = 0$.

Let $\widehat{\gamma}$ be the 2SLS or IV estimator of γ in model (3),

$$\widehat{\gamma} = \left(Z_2' M_{\widehat{X}} Z_2\right)^{-1} Z_2' M_{\widehat{X}} y. \tag{5}$$

It then follows from Proposition 1 that the Hansen test for overidentifying restrictions is equal to the robust score test

$$J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) = \widehat{\gamma}' \left(V\widehat{a}r_{r,\widehat{u}_1}\left(\widehat{\gamma}\right)\right)^{-1} \widehat{\gamma},$$

where $V\widehat{a}r_{r,\widehat{u}_1}\left(\widehat{\gamma}\right)$ is a robust estimator of the variance of $\widehat{\gamma}$ under the null,

$$V\widehat{a}r_{r,\widehat{u}_1}\left(\widehat{\gamma}\right) = \left(Z_2' M_{\widehat{X}} Z_2\right)^{-1} \left(Z_2' M_{\widehat{X}} H_{\widehat{u}_1} M_{\widehat{X}} Z_2\right)^{-1} \left(Z_2' M_{\widehat{X}} Z_2\right)^{-1}.$$

Equivalently, the non-robust score test $S\left(\widehat{\beta}_1\right)$ can be obtained by using the non-robust variance estimator $V\widehat{a}r_{\widehat{u}_1}\left(\widehat{\gamma}\right) = \frac{\widehat{u}_1'\widehat{u}_1}{n}\left(Z_2' M_{\widehat{X}} Z_2\right)^{-1}$. These versions of the test can alternatively be obtained as a test for $H_0 : \gamma = 0$ after OLS regression of the specification

$$\widehat{u}_1 = \widehat{X}\eta + Z_2\gamma + \xi_1, \tag{6}$$

and estimating the (robust) variance of $\widehat{\gamma}$ under the null. Wooldridge (1995) considered this score test approach for the 2SLS estimator, but did not establish its equivalence with the Hansen J-test, see also Baum, Schaffer and Stillman (2007).

Alternatively, one could base the (robust) variance estimator of $\widehat{\gamma}$ in (6) on the residuals

$$\widehat{\xi}_1 = M_Z \widehat{u}_1 = \widehat{v}_y - \widehat{V}\widehat{\beta}_1,$$

where $\widehat{v}_y = M_Z y$ and $\widehat{V} = M_Z X$ are the first-stage OLS residuals in the linear projection specifications

$$y = Z\pi_y + v_y, \tag{7}$$

and (2). This leads to the robust score test

$$B_r\left(\widehat{\beta}_1\right) = y' M_{\widehat{X}} Z_2 \left(Z_2' M_{\widehat{X}} H_{\widehat{\xi}_1} M_{\widehat{X}} Z_2\right)^{-1} Z_2' M_{\widehat{X}} y,$$

which is equivalent to the Hansen J-test based on the two-step GMM estimator

$$\widehat{\beta}_{2,\widehat{\xi}_1} = \left(X'Z\left(Z'H_{\widehat{\xi}_1}Z\right)^{-1}Z'X\right)^{-1} X'Z\left(Z'H_{\widehat{\xi}_1}Z\right)^{-1}Z'y.$$

The versions $S_r\left(\widehat{\beta}_1\right)$ and $B_r\left(\widehat{\beta}_1\right)$ can be characterised as robust Sargan and Basmann score tests, as per the proposals by Sargan (1958) and Basmann (1960) for the LIML estimator as discussed below. As they can be seen to be the score and Wald tests respectively for the null $H_0 : \gamma = 0$ in specification (6), they are sometimes referred to as the LM and Wald versions of the test, see e.g. Baum, Schaffer and Stillman (2007) and Bazzi and Clemens (2013) in their discussions of the Cragg-Donald and Kleibergen-Paap rank tests. This is perhaps confusing as both versions are score tests for the null $H_0 : \gamma = 0$ in model (3), but with the variance based on different versions of the residual under the null. From (1), (7) and (2), it follows that $\pi_y = \Pi\beta$ and $u = v_y - V\beta$, and so both $\widehat{u}_1$ and $\widehat{\xi}_1$ are proxies for the same error.

Let $Z = \left[\begin{array}{cc} Z_1 & Z_2 \end{array}\right]$, with Z_1 an $n \times k_x$ matrix. Partition π_y and Π accordingly as $\pi_y = \left[\begin{array}{cc} \pi_{y1}' & \pi_{y2}' \end{array}\right]'$ and $\Pi = \left[\begin{array}{cc} \Pi_1' & \Pi_2' \end{array}\right]'$. Assume that Π_1 has full rank k_x. From the linear projections (7) and (2), we have that

$$\begin{aligned} y &= Z_1\pi_{y1} + Z_2\pi_{y2} + v_y \\ &= Z_1\Pi_1\Pi_1^{-1}\pi_{y1} + Z_2\pi_{y2} + v_y \\ &= X\Pi_1^{-1}\pi_{y1} + Z_2\left(\pi_{y2} - \Pi_2\Pi_1^{-1}\pi_{y_1}\right) + v_y - V\Pi_1^{-1}\pi_{y1} \\ &= X\kappa + Z_2\gamma + w, \end{aligned} \tag{8}$$

where $\kappa = \Pi_1^{-1}\pi_{y1}$. It follows that if $\pi_y = \Pi\beta$, then $\kappa = \beta$ and $\gamma = 0$. Consider the IV estimator for γ as in (5). Let $Z^* = \begin{bmatrix} \widehat{X} & Z_2 \end{bmatrix} = Z\widehat{D}$, with

$$\widehat{D} = \begin{bmatrix} \widehat{\Pi}_1 & 0 \\ \widehat{\Pi}_2 & I_{k_z-k_x} \end{bmatrix}; \quad \widehat{D}^{-1} = \begin{bmatrix} \widehat{\Pi}_1^{-1} & 0 \\ -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix},$$

then the IV estimator of $\theta = \begin{pmatrix} \beta' & \gamma' \end{pmatrix}'$ in (3) is given by

$$\begin{aligned} \widehat{\theta} &= \left(Z^{*\prime}Z^*\right)^{-1} Z^{*\prime}y = \widehat{D}^{-1}\left(Z'Z\right)^{-1} Z'y \\ &= \widehat{D}^{-1}\widehat{\pi}_y. \end{aligned}$$

Hence $\widehat{\beta} = \widehat{\Pi}_1^{-1}\widehat{\pi}_{y1}$ and

$$\widehat{\gamma} = \widehat{\pi}_{y2} - \widehat{\Pi}_2\widehat{\Pi}_1^{-1}\widehat{\pi}_{y_1} = \widehat{\pi}_{y2} - \widehat{\Pi}_2\widehat{\beta}.$$

For later reference, note also that the OLS estimator for the parameters in (6) is given by $\widehat{D}^{-1}\left(Z'Z\right)^{-1} Z'\widehat{u}_1 = \widehat{D}^{-1}\widehat{\Pi}^*\widehat{\psi}_1$, where $\widehat{\psi}_1 = \begin{pmatrix} 1 & -\widehat{\beta}_1 \end{pmatrix}$, and $\widehat{\Pi}^* = \left(Z'Z\right)^{-1} Z'W$ is the OLS estimator of $\Pi^* = \begin{bmatrix} \pi_y & \Pi \end{bmatrix}$, with $W = \begin{bmatrix} y & X \end{bmatrix}$. Hence $\widehat{\gamma}$ can also be expressed as

$$\widehat{\gamma} = \begin{bmatrix} -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix} \widehat{\Pi}^*\widehat{\psi}_1. \tag{9}$$

3.1 LIML and CU-GMM

Whilst the above score tests remain valid also for the LIML estimator $\widehat{\beta}_L$ of β, treating it as a one-step GMM estimator, the test statistics using the projection $\widehat{X}$ are not invariant to normalisation, i.e. the choice of endogenous variable as the dependent variable. The LIML estimator estimates both β and the first stage parameters Π. Whilst these are equal to the 2SLS and OLS estimators respectively in the just-identified unrestricted model (3), they are equal to $\widehat{\beta}_L$ and $\widehat{\Pi}_L$ in the restricted model.

Let $\widehat{\Sigma}_w = \frac{1}{n}W'W$, $\widehat{V}^* = \begin{bmatrix} \widehat{v}_y & \widehat{V} \end{bmatrix}$ and $\widehat{\Sigma}_{v^*} = \frac{1}{n}\widehat{V}^{*\prime}\widehat{V}^*$. Following Alonso-Borrego and Arellano (1999) the LIML estimator for β can be obtained as the Continuous Updating Estimator,

$$\widehat{\beta}_L = \arg\min_{\beta} S\left(\beta\right) = \arg\min_{\beta} B\left(\beta\right), \tag{10}$$

where

$$\begin{aligned}
S(\beta) &= \frac{(y - X\beta)' P_Z (y - X\beta)}{(y - X\beta)'(y - X\beta)/n} = \frac{\psi' W' P_Z W' \psi}{\psi' \widehat{\Sigma}_w \psi} \\
B(\beta) &= \frac{(y - X\beta)' P_Z (y - X\beta)}{\left(\widehat{v}_y - \widehat{V}\beta\right)' \left(\widehat{v}_y - \widehat{V}\beta\right)/n} = \frac{\psi' W' P_Z W' \psi}{\psi' \widehat{\Sigma}_{v^*} \psi}
\end{aligned}$$

and the second equality in (10) is easily shown to hold as $\widehat{V}^* = M_Z W$.

The LIML estimator of ψ is therefore equal to $\widehat{\Sigma}_w^{-1/2} v_{[1]}$ where $v_{[1]}$ is the eigenvector associated with $e_{[1]}$, the minimum eigenvalue of $\widehat{\Sigma}_w^{-1/2} W' P_Z W \widehat{\Sigma}_w^{-1/2}$, and

$$S\left(\widehat{\beta}_L\right) = e_{[1]}. \tag{11}$$

Denote $\pi^* = \text{vec}(\Pi^*)$ and $\widehat{\pi}^* = \text{vec}\left(\widehat{\Pi}^*\right)$. Under the standard assumptions we have that

$$\sqrt{n}\left(\widehat{\pi}^* - \pi^*\right) \xrightarrow{d} N\left(0, V_{\pi^*}\right),$$

with, in the homoskedastic case, $V_{\pi^*} = \Sigma_{v^*} \otimes Q_{zz}^{-1}$. An estimator for the variance of $\widehat{\pi}^*$ is therefore $V\widehat{a}r_{v^*}(\widehat{\pi}^*) = \widehat{\Sigma}_{v^*} \otimes (Z'Z)^{-1}$. Let $\pi = \text{vec}(\Pi)$ and $\widehat{\pi} = \text{vec}\left(\widehat{\Pi}\right)$, then the LIML estimator for β and Π can be obtained as the minimum distance estimator

$$\begin{aligned}
\left(\widehat{\beta}_L, \widehat{\Pi}_L\right) &= \arg\min_{\beta,\Pi} MD_v(\beta, \Pi); \\
MD_v(\beta, \Pi) &= \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(V\widehat{a}r_{v^*}(\widehat{\pi}^*)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix} \\
&= \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(\widehat{\Sigma}_{v^*}^{-1} \otimes (Z'Z)\right) \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix},
\end{aligned} \tag{12}$$

and $MD_v\left(\widehat{\beta}_L, \widehat{\Pi}_L\right) = B\left(\widehat{\beta}_L\right)$, see Alonso-Borrego and Arellano (1999).

Equivalently, if we specify the variance estimator as $V\widehat{a}r_w(\widehat{\pi}^*) = \widehat{\Sigma}_w^{-1} \otimes (Z'Z)$ instead we obtain

$$\begin{aligned}
\left(\widehat{\beta}_L, \widehat{\Pi}_L\right) &= \arg\min_{\beta,\Pi} MD_w(\beta, \Pi); \\
MD_w(\beta, \Pi) &= \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(\widehat{\Sigma}_w^{-1} \otimes (Z'Z)\right) \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}
\end{aligned}$$

and $MD_w\left(\widehat{\beta}_L, \widehat{\Pi}_L\right) = S\left(\widehat{\beta}_L\right)$. Note that $V\widehat{a}r_w(\widehat{\pi}^*)$ is an estimator of the variance of $\widehat{\pi}^*$ under the null that $\Pi^* = 0$. Although this is not a valid restriction, it results in using $\widehat{u}_L = y - X\widehat{\beta}_L$ as the residual, whereas the use of $V\widehat{a}r_{v^*}(\widehat{\pi}^*)$ results in the residual $\widehat{\xi}_L = M_Z\widehat{u}_L = \widehat{v}_y - \widehat{V}\widehat{\beta}_L$.

We have the following general result for a class of minimum distance estimators that includes LIML.

Lemma 1 *Consider the minimum distance estimators*

$$\left(\widehat{\beta}_A, \widehat{\Pi}_A\right) = \arg\min_{\beta,\Pi} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' (A \otimes (Z'Z)) \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix},$$

with A a symmetric nonsingular $k_x \times k_x$ matrix. Let $\widehat{X}_A = Z\widehat{\Pi}_A$, then

$$\begin{aligned} \widehat{\beta}_A &= \left(\widehat{X}_A'X\right)^{-1}\widehat{X}_A'y \\ &= \left(\widehat{X}_A'\widehat{X}_A\right)^{-1}\widehat{X}_A'y. \end{aligned}$$

Proof. See Appendix ■

For the LIML estimator, let $\widehat{X}_L = Z\widehat{\Pi}_L$. It is well known that $\widehat{\beta}_L = \left(\widehat{X}_L'X\right)^{-1}\widehat{X}_L'y$, see e.g. Bowden and Turkington (1984, p.113), but it follows from Lemma 1 that we also have that $\widehat{\beta}_L = \left(\widehat{X}_L'\widehat{X}_L\right)^{-1}\widehat{X}_L'y$. Let $\widehat{u}_L = y - X\widehat{\beta}_L$ and $\widehat{C}_L = \begin{bmatrix} \widehat{\beta}_L & I_{k_x} \end{bmatrix}' \otimes I_{k_z}$. From the minimum distance first-order condition we get

$$\text{vec}\left(\widehat{\Pi}_L\right) = \left(\widehat{C}_L'\left(V\widehat{a}r\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{C}_L\right)^{-1}\widehat{C}_L'\left(V\widehat{a}r\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{\pi}^*. \tag{13}$$

As Godfrey and Wickens (1982) show, $\widehat{\Pi}_L$ can equivalently be calculated as

$$\widehat{\Pi}_L = \left(Z'M_{\widehat{u}_L}Z\right)^{-1}Z'M_{\widehat{u}_L}X.$$

It then follows straightforwardly that the robust score test for overidentifying restrictions based on the LIML estimators for Π and β in the restricted model, and which is invariant to normalisation, is given by

$$S_r\left(\widehat{\beta}_L\right) = \widehat{u}_L'M_{\widehat{X}_L}Z_2\left(Z_2'M_{\widehat{X}_L}H_{\widehat{u}_L}M_{\widehat{X}_L}Z_2\right)^{-1}Z_2'M_{\widehat{X}_L}\widehat{u}_L. \tag{14}$$

As $\widehat{\beta}_L = \left(\widehat{X}_L'\widehat{X}_L\right)^{-1}\widehat{X}_L'y$, we get from standard score test theory for the non-robust version of the test, as above for the 2SLS estimator,

$$S\left(\widehat{\beta}_L\right) = \frac{\widehat{u}_L'P_Z\widehat{u}_L}{\widehat{u}_L'\widehat{u}_L/n},$$

which is Sargan's (1958) test of overidentifying restrictions.

Let $\widehat{\xi}_L = \widehat{v}_y - \widehat{V}\widehat{\beta}_L$, then the Basmann (1960) version of the overidentification test is given by

$$B\left(\widehat{\beta}_L\right) = \frac{\widehat{u}_L' P_Z \widehat{u}_L}{\widehat{\xi}_L' \widehat{\xi}_L / n},$$

with robust version

$$B_r\left(\widehat{\beta}_L\right) = \widehat{u}_L' M_{\widehat{X}_L} Z_2 \left(Z_2' M_{\widehat{X}_L} H_{\widehat{\xi}_L} M_{\widehat{X}_L} Z_2\right)^{-1} Z_2' M_{\widehat{X}_L} \widehat{u}_L. \quad (15)$$

As for the GMM analysis above, we can obtain the LIML based test statistics as (robust) tests for $H_0 : \gamma = 0$ after OLS estimation of the specification

$$\widehat{u}_L = \widehat{X}_L \eta + Z_2 \gamma + \xi_L.$$

Another robust invariant test of overidentifying restrictions is the Hansen J-test $J\left(\widehat{\beta}_{cu}\right)$, based on the continuously updated CU-GMM estimator

$$\begin{aligned} \widehat{\beta}_{cu} &= \arg\min_{\beta} J(\beta); \\ J(\beta) &= (y - X\beta)' Z \left(Z' H_{u(\beta)} Z\right)^{-1} Z' (y - X\beta), \end{aligned}$$

where $u(\beta) = y - X\beta$, and $J\left(\widehat{\beta}_{cu}\right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under Assumptions 1-5.

The Basmann CU-GMM version is given by $J_v\left(\widehat{\beta}_{cu,v}\right)$, obtained as

$$\begin{aligned} \widehat{\beta}_{cu,v} &= \arg\min_{\beta} J_v(\beta); \\ J_v(\beta) &= (y - X\beta)' Z \left(Z' H_{\xi(\beta)} Z\right)^{-1} Z' (y - X\beta), \end{aligned}$$

with $\xi(\beta) = \widehat{v}_y - \widehat{V}\beta$.

If we specify a robust estimator for the variance of $\widehat{\pi}^*$ on the basis of the OLS residuals $\widehat{V}^*$, denoted $V\widehat{a}r_{r,v^*}(\widehat{\pi}^*)$, then Kleibergen and Mavroeidis (2008, Appendix) show that

$$\begin{aligned} \left(\widehat{\beta}_{cu,v}, \widehat{\Pi}_{cu,v}\right) &= \arg\min_{\beta,\Pi} MD_{v,r}(\beta, \Pi); \\ MD_{v,r}(\beta, \Pi) &= \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(V\widehat{a}r_{r,v^*}(\widehat{\pi}^*)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix} \end{aligned}$$

and

$$J_v\left(\widehat{\beta}_{cu,v}\right) = MD_{v,r}\left(\widehat{\beta}_{cu,v}, \widehat{\Pi}_{cu,v}\right),$$

see also Gospodinov, Kan and Robotti (2017). Alternatively, specifying the robust estimator of the variance of $\widehat{\pi}^*$ under the null that $\Pi^* = 0$ and denoted $V\widehat{a}r_{r,w}(\widehat{\pi}^*)$, results in

$$\begin{aligned}
\left(\widehat{\beta}_{cu}, \widehat{\Pi}_{cu}\right) &= \arg\min_{\beta,\Pi} MD_r(\beta,\Pi); \\
MD_r(\beta,\Pi) &= \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(V\widehat{a}r_{r,w}(\widehat{\pi}^*)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix} \\
J\left(\widehat{\beta}_{cu}\right) &= MD_r\left(\widehat{\beta}_{cu}, \widehat{\Pi}_{cu}\right).
\end{aligned}$$

$J\left(\widehat{\beta}_{cu}\right)$ is equal to the robust score test evaluated at the CU-GMM estimates

$$J\left(\widehat{\beta}_{cu}\right) = S_r\left(\widehat{\beta}_{cu}\right) = \widehat{u}'_{cu} M_{\widehat{X}_{cu}} Z_2 \left(Z_2' M_{\widehat{X}_{cu}} H_{\widehat{u}_{cu}} M_{\widehat{X}_{cu}} Z_2\right)^{-1} Z_2' M_{\widehat{X}_{cu}} \widehat{u}_{cu},$$

where $\widehat{u}_{cu} = y - X\widehat{\beta}_{cu}$and $\widehat{X}_{cu} = Z\widehat{\Pi}_{cu}$.

The standard CU-GMM estimator in the literature is $\widehat{\beta}_{cu}$. Use of the alternative estimator $\widehat{\beta}_{cu,v}$ is less common. However, Chernozhukov and Hansen (2008) propose a method for weak instrument robust inference as follows. For a sequence of values $b \in B$, test the null $H_0 : \alpha = 0$ in the regression model

$$y - Xb = Z\alpha + \varepsilon$$

by a (robust) Wald test, denoted $W(b)$. Then construct the $1-p$ confidence regions as the set of b such that $W(b) \leq c(1-p)$, where $c(1-p)$ is the $(1-p)^{th}$ percentile of the $\chi^2_{k_z}$ distribution, see Chernozhukov and Hansen (2008, p. 69). As the Wald test uses the residuals $M_Z\varepsilon = \widehat{v}_y - \widehat{V}b = \xi(b)$ it is clear that this procedure is equivalent to finding the values b such that

$$J_v(b) = (y - Xb)' Z \left(Z' H_{\xi(b)} Z\right)^{-1} Z'(y - Xb) \leq c(1-p).$$

The LIML and CU-GMM based score tests achieve their invariance properties by changing the estimator for γ in (3) as compared to the IV estimator (5). As in (9), we get for the LIML estimator that

$$\begin{aligned}
\widehat{\gamma}_L &= \begin{bmatrix} -\widehat{\Pi}_{2L}\widehat{\Pi}_{1L}^{-1} & I_{k_z-k_x} \end{bmatrix} \widehat{\Pi}^* \widehat{\psi}_L \\
&= \left(\widehat{\psi}_L' \otimes \begin{bmatrix} -\widehat{\Pi}_{2L}\widehat{\Pi}_{1L}^{-1} & I_{k_z-k_x} \end{bmatrix}\right) \text{vec}\left(\widehat{\Pi}^*\right).
\end{aligned} \tag{16}$$

Equivalently, for the CU-GMM estimator, we get

$$\widehat{\gamma}_{cu} = \left(\widehat{\psi}_{cu}' \otimes \begin{bmatrix} -\widehat{\Pi}_{2,cu}\widehat{\Pi}_{1,cu}^{-1} & I_{k_z-k_x} \end{bmatrix}\right) \text{vec}\left(\widehat{\Pi}^*\right).$$

3.2 Relationship to CD and KP Rank Tests

The test for overidentifying restrictions is equivalent to testing whether the rank of the $k_z \times (k_x + 1)$ matrix Π^* is equal k_x. The Cragg and Donald (1993, 1997) rank test is defined as

$$CD = \min_{\pi^*} \left(\widehat{\pi}^* - \pi^*\right) \left(V\widehat{a}r\left(\widehat{\pi}^*\right)\right)^{-1\prime} \left(\widehat{\pi}^* - \pi^*\right);$$

$$\text{s.t. } \mathrm{r}\left(\Pi^*\right) = k_x,$$

or equivalently,

$$CD = \min_{\Pi,\beta} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(V\widehat{a}r\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}.$$

CD is therefore equal to the minimum distance criterion $MD_v\left(\widehat{\beta}_L, \widehat{\Pi}_L\right) = B\left(\widehat{\beta}_L\right)$, when the variance estimator is specified as the non-robust $V\widehat{a}r_{v^*}\left(\widehat{\pi}^*\right)$. The CD statistic is equal to $S\left(\widehat{\beta}_L\right)$ when specifying it as $V\widehat{a}r_w\left(\widehat{\pi}^*\right)$. Specifying robust variance estimators results in the CD statistic to be equal to either the CU-GMM tests $J_v\left(\widehat{\beta}_{cu,v}\right)$ or $J\left(\widehat{\beta}_{cu}\right)$.

For the Kleibergen and Paap (2006) rank test, let G and F be $k_z \times k_z$ and $(k_x + 1) \times (k_x + 1)$ finite non-singular matrices respectively, and define

$$\Theta = G\Pi^* F'; \quad \widehat{\Theta} = G\widehat{\Pi}^* F'.$$

For testing $H_0 : \mathrm{r}\left(\Pi^*\right) = q$, Kleibergen-Paap (KP) propose use of the singular value decomposition (SVD)

$$\Theta = USU^{*\prime},$$

where U and U^* are $k_z \times k_z$ and $(k_x + 1) \times (k_x + 1)$ orthonormal matrices respectively, and S is a $k_z \times (k_x + 1)$ matrix that contains the singular values of Θ on its main diagonal and is equal to zero elsewhere. KP show that the SVD results in the decomposition of Θ as

$$\Theta = A_q B_q + A_{q,\perp} \Lambda_q B_{q,\perp},$$

where A_q is a $k_z \times q$ matrix, B_q is a $q \times (k_x + 1)$ matrix, $A_{q,\perp}$ is a $k_z \times (k_z - q)$ matrix, Λ_q is $(k_z - q)\left((k_x + 1) - q\right)$ matrix, $B_{q,\perp}$ is a $\left((k_x + 1) - q\right) \times (k_x + 1)$ matrix; $A_q' A_{q,\perp} = 0$ and $B_{q,\perp} B_q' = 0$. As $\Lambda_q = 0$ under the null $H_0 : \mathrm{r}\left(\Pi^*\right) = q$, the KP test is a test for $H_0 : \mathrm{vec}\left(\Lambda_q\right) = 0$.

The SVD applied to $\widehat{\Theta}$ yields the decomposition,

$$\begin{aligned}\widehat{\Theta} &= \widehat{A}_q \widehat{B}_q + \widehat{A}_{q,\perp} \widehat{\Lambda}_q \widehat{B}_{q,\perp};\\ \widehat{\Lambda}_q &= \widehat{A}'_{q,\perp} \widehat{\Theta} \widehat{B}'_{q,\perp},\end{aligned}$$

and the KP test statistic is given by

$$\mathrm{rk}\,(q) = \widehat{\lambda}'_q \widehat{\Omega}_q^{-1} \widehat{\lambda}_q, \tag{17}$$

where $\widehat{\lambda}_q = \mathrm{vec}\left(\widehat{\Lambda}_q\right) = \left(\widehat{B}_{q,\perp} \otimes \widehat{A}'_{q,\perp}\right) \mathrm{vec}\left(\widehat{\Theta}\right) = \left(\widehat{B}_{q,\perp} F \otimes \widehat{A}'_{q,\perp} G\right) \widehat{\pi}^*$; and $\widehat{\Omega}_q$ is an estimator of the asymptotic variance of $\widehat{\lambda}_q$. Robust versions of the test are obtained by specifying a robust estimator of the variance of $\widehat{\pi}^*$.

As $\mathrm{vec}\left(\widehat{\Theta}\right) = (F \otimes G)\, \widehat{\pi}^*$ it follows that $Var\left(\mathrm{vec}\left(\widehat{\Theta}\right)\right) = (F \otimes G)\, Var\,(\widehat{\pi}^*)\,(F' \otimes G')$ and KP argue that it is best to specify F and G such that $(F \otimes G)\, V_{\pi^*}\,(F' \otimes G')$ is close to the identity matrix. Hence, assuming homoskedasticity, one would choose F and G such that $F'F = \widehat{\Sigma}_{v^*}^{-1}$ and $G'G = \widehat{\Sigma}_z = Z'Z/n$.

The next Proposition gives the relationship between the Kleibergen-Paap test and the robust score test.

Proposition 2 *Consider the Kleibergen-Paap rank test (17) for* $H_0 : \mathrm{r}\,(\Pi^*) = k_x$*. Given choices of* F *and* G*, define the estimators* $\widehat{\beta}_{GF}$ *and* $\widehat{\Pi}_{GF}$ *as*

$$\left(\widehat{\beta}_{GF}, \widehat{\Pi}_{GF}\right) = \arg\min_{\beta,\Pi} \left(\begin{pmatrix} \widehat{\pi}_y \\ \widehat{\pi} \end{pmatrix} - \begin{pmatrix} \Pi\beta \\ \pi \end{pmatrix}\right)' (F'F \otimes G'G) \left(\begin{pmatrix} \widehat{\pi}_y \\ \widehat{\pi} \end{pmatrix} - \begin{pmatrix} \Pi\beta \\ \pi \end{pmatrix}\right).$$

Let $\widehat{u}_{GF} = y - X\widehat{\beta}_{GF}$*,* $\widehat{X}_{GF} = Z\widehat{\Pi}_{GF}$ *and let* Z_2 *be a matrix of any* $k_z - k_x$ *subset of instruments. Then*

$$\mathrm{rk}\,(k_x) = \widehat{u}'_{GF} M_{\widehat{X}_{GF}} Z_2 \left(Z_2' M_{\widehat{X}_{GF}} H_{\widehat{r}} M_{\widehat{X}_{GF}} Z_2\right)^{-1} Z_2' M_{\widehat{X}_{GF}} \widehat{u}_{GF},$$

where the residual $\widehat{r}$ *is either equal to* $\widehat{u}_{GF}$ *or equal to* $\widehat{\xi}_{GF} = \widehat{v}_y - \widehat{V}\widehat{\beta}_{GF}$*. This choice of residual and the robustness of the test is determined by the choice of the estimator of* $Var\,(\widehat{\pi}^*)$ *used in the estimator* $\widehat{\Omega}_q$ *of* $Var\left(\widehat{\lambda}_q\right)$*.*

Proof. See Appendix ■

It follows from Proposition 2 that the various versions of the KP test can be obtained as the tests for $H_0 : \gamma = 0$ in the specification

$$\widehat{u}_{GF} = \widehat{X}_{GF}\eta + Z_2\gamma + \xi_{GF},$$

estimated by OLS.

The estimator $\widehat{\beta}_{GF}$ can alternatively be obtained as the continuous updating estimator

$$\widehat{\beta}_{GF} = \arg\min_{\beta} \frac{(y - X\beta)' Z (Z'Z)^{-1} G'G (Z'Z)^{-1} Z' (y - X\beta)}{\begin{pmatrix} 1 & -\beta' \end{pmatrix}' (F'F)^{-1} \begin{pmatrix} 1 & -\beta' \end{pmatrix}'}. \tag{18}$$

It is clear that choosing F and G such that $F'F = \widehat{\Sigma}_{v^*}^{-1}$ or $F'F = \widehat{\Sigma}_{w}^{-1}$ and $G'G = \widehat{\Sigma}_z$ results in the LIML estimators for β and Π. Choosing alternatively $F'F = I_{k_x}$ and $G'G = \widehat{\Sigma}_z$ results in the symmetrically normalised 2SLS estimator, see Alonso-Borrego and Arellano (1999).

The robust KP tests commonly reported in standard estimation routines, like ivreg2 in Stata (Baum, Schaffer and Stillman, 2010), are based on the LIML normalisation. It therefore follows that these LIML based versions of $\text{rk}(k_x)$ are equal to $S_r\left(\widehat{\beta}_L\right)$ or $B_r\left(\widehat{\beta}_L\right)$ as defined in (14) and (15) respectively, depending on the choice of robust estimator of $Var(\widehat{\pi}^*)$.

3.3 A Two-Step Invariant Estimator and Test for Overidentifying Restrictions

If Π were known, then the natural just-identifying linear combination of instrument would be $\widetilde{Z} = Z\Pi$, which would be the efficient combination in the homoskedastic model. 2SLS and LIML are asymptotically efficient in that case by estimating Π consistently by $\widehat{\Pi}$ and $\widehat{\Pi}_L$ respectively.

For a general known Ω_{zu}, the optimal combination of instruments for known Π is

$$\widetilde{Z} = Z'\Omega_{zu}^{-1} Z'Z\Pi$$

and the efficient IV estimator is given by

$$\begin{aligned} \widehat{\beta} &= \left(\widetilde{Z}'X\right)^{-1} \widetilde{Z}'y \\ &= \left(\Pi'Z'Z\Omega_{zu}^{-1}Z'X\right)^{-1} \Pi'Z'Z\Omega_{zu}^{-1}Z'y, \end{aligned}$$

with limiting distribution

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \left(\Pi' Q_{zz} \Omega_{zu}^{-1} Q_{zz} \Pi\right)^{-1}\right). \tag{19}$$

For the 2SLS estimator and associated two-step GMM estimator, Π is estimated by OLS, $\widehat{\Pi} = (Z'Z)^{-1} Z'X$, and

$$\widehat{\beta}_{2sls} = \left(\widehat{\Pi}' Z'X\right)^{-1} \widehat{\Pi}' Z'y.$$

Letting $\widehat{u}_{2sls} = y - X\widehat{\beta}_{2sls}$, the two-step GMM estimator is given by

$$\begin{aligned}\widehat{\beta}_2 &= \left(\widehat{\Pi}' Z'Z \left(Z'H_{\widehat{u}_{2sls}} Z\right)^{-1} Z'X\right)^{-1} \widehat{\Pi}' Z'Z \left(Z'H_{\widehat{u}_{2sls}} Z\right)^{-1} Z'y \\ &= \left(X'Z \left(Z'H_{\widehat{u}_{2sls}} Z\right)^{-1} Z'X\right)^{-1} X'Z \left(Z'H_{\widehat{u}_{2sls}} Z\right)^{-1} Z'y,\end{aligned}$$

which is asymptotically efficient with the same limiting distribution as the infeasible estimator (19).

For the LIML estimator of β, we have

$$\widehat{\beta}_L = \left(\widehat{\Pi}_L' Z'X\right)^{-1} \widehat{\Pi}_L' Z'y,$$

with $\widehat{\Pi}_L$ the LIML estimator of Π. An optimal invariant two-step estimator is then given by

$$\widehat{\beta}_{2L} = \left(\widehat{\Pi}_L' Z'Z \left(Z'H_{\widehat{u}_L} Z\right)^{-1} Z'X\right)^{-1} \widehat{\Pi}_L' Z'Z \left(Z'H_{\widehat{u}_L} Z\right)^{-1} Z'y, \tag{20}$$

with $\widehat{u}_L = y - X\widehat{\beta}_L$. $\widehat{\beta}_{2L}$ has the same limiting distribution as the infeasible estimator, and, like LIML, is invariant to normalisation. The Hansen J-statistic calculated as

$$J\left(\widehat{\beta}_{2L}\right) = \widehat{u}_{2L}' Z \left(Z'H_{\widehat{u}_{2L}} Z\right)^{-1} Z'\widehat{u}_{2L}, \tag{21}$$

with $\widehat{u}_{2L} = y - X\widehat{\beta}_{2L}$ is also invariant to normalisation.

3.4 CU-GMM as an Iterated GMM Estimator

For the CU-GMM estimator we have the following result. For brevity we focus on $\widehat{\beta}_{cu}$ and $\widehat{\Pi}_{cu}$, but results straightforwardly carry over to $\widehat{\beta}_{cu,v}$ and $\widehat{\Pi}_{cu,v}$.

Lemma 2 *Consider the CU-GMM minimum distance estimators*

$$\left(\widehat{\beta}_{cu}, \widehat{\Pi}_{cu}\right) = \arg\min \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix}' \left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi\beta \\ \widehat{\pi} - \pi \end{pmatrix},$$

and let $\widehat{u}_{cu} = y - X\widehat{\beta}_{cu}$. Then

$$\widehat{\beta}_{cu} = \left(\widehat{\Pi}_{cu}' Z'Z \left(Z'H_{\widehat{u}_{cu}} Z\right)^{-1} Z'X\right)^{-1} \widehat{\Pi}_{cu}' Z'Z \left(Z'H_{\widehat{u}_{cu}} Z\right)^{-1} Z'y.$$

Proof. See Appendix ■

From Lemma 2 it is clear that the main difference between the two-step GMM estimator and the CU-GMM estimator is the estimator for Π, with the two-step GMM estimator keeping this fixed at the OLS estimator $\widehat{\Pi}$.

Similar to the first-order condition (13) for the LIML estimator of Π, we have for $\widehat{\Pi}_{cu}$

$$\text{vec}\left(\widehat{\Pi}_{cu}\right) = \left(\widehat{C}'_{cu}\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{C}_{cu}\right)^{-1}\widehat{C}'_{cu}\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{\pi}^*, \tag{22}$$

where $\widehat{C}_{cu} = \left[\begin{array}{cc}\widehat{\beta}_{cu} & I_{k_x}\end{array}\right]' \otimes I_{k_z}$.

Let $\widehat{\beta}_1$ be an initial consistent and normal estimator for β. Let $\widehat{C}_1 = \left[\begin{array}{cc}\widehat{\beta}_1 & I_{k_x}\end{array}\right]' \otimes I_{k_z}$, and

$$\text{vec}\left(\widehat{\Pi}_1\right) = \left(\widehat{C}'_1\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{C}_1\right)^{-1}\widehat{C}'_1\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{\pi}^*.$$

Then an alternative two-step GMM estimator is given by

$$\widehat{\beta}_2 = \left(\widehat{\Pi}'_1 Z'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'X\right)^{-1}\widehat{\Pi}'_1 Z'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'y,$$

and a general iteration scheme then is

$$\widehat{\beta}_{j+1} = \left(\widehat{\Pi}'_j Z'Z\left(Z'H_{\widehat{u}_j}Z\right)^{-1}Z'X\right)^{-1}\widehat{\Pi}'_j Z'Z\left(Z'H_{\widehat{u}_j}Z\right)^{-1}Z'y,$$

resulting in $\widehat{\beta}_{cu}$ and $\widehat{\Pi}_{cu}$ upon convergence.

It is interesting to note that if $\widehat{\beta}_1$ is not an invariant estimator, we obtain a sequence of efficient estimators converging to an invariant estimator. If $\widehat{\beta}_1$ is an invariant estimator, for example the LIML estimator $\widehat{\beta}_L$, we obtain a sequence of efficient invariant estimators. Note that therefore an alternative invariant two-step estimator to $\widehat{\beta}_{2L}$ in (20) is given by

$$\widehat{\beta}_{2L,r} = \left(\widehat{\Pi}'_{L,r} Z'Z\left(Z'H_{\widehat{u}_L}Z\right)^{-1}Z'X\right)^{-1}\widehat{\Pi}'_{L,r} Z'Z\left(Z'H_{\widehat{u}_L}Z\right)^{-1}Z'y,$$

where

$$\text{vec}\left(\widehat{\Pi}_{L,r}\right) = \left(\widehat{C}'_L\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{C}_L\right)^{-1}\widehat{C}'_L\left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1}\widehat{\pi}^*,$$

with invariant Hansen test

$$J\left(\widehat{\beta}_{2L,r}\right) = \widehat{u}'_{2L,r}Z\left(Z'H_{\widehat{u}_{2L,r}}Z\right)^{-1}Z'\widehat{u}_{2L,r},$$

where $\widehat{u}_{2L,r} = y - X\widehat{\beta}_{2L,r}$.

4 Underidentification Tests

Assumption 5, $E(z_i x_i') = Q_{zx}$ has full column rank k_x, is a necessary condition for the identification of β using the instrumental variables z_i. As $\Pi = Q_{zz}^{-1} E(z_i x_i')$ it follows that the rank of Π is equal to the rank of $E(z_i x_i')$. This means that β is identified iff $\mathrm{r}(\Pi) = k_x$.

Standard tests for underidentification, like the Cragg-Donald and Kleibergen-Paap tests are tests for $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$. If $\mathrm{r}(\Pi) = k_x - 1$, then there is a k_x-vector δ^*, such that $\Pi\delta^* = 0$. Partition $X = \begin{bmatrix} x_1 & X_2 \end{bmatrix}$, with x_1 an n-vector and X_2 an $n \times (k_x - 1)$ matrix, and equivalently $V = \begin{bmatrix} v_1 & V_2 \end{bmatrix}$, $\Pi = \begin{bmatrix} \pi_1 & \Pi_2 \end{bmatrix}$ and $\delta^* = \begin{bmatrix} \delta_1^* & \delta_2^{*\prime} \end{bmatrix}'$. Assuming $\delta_1^* \neq 0$, then $\pi_1 = \Pi_2 \delta$, with $\delta = -\delta_2^* / \delta_1^*$ and hence

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 = Z\Pi_2\delta + v_1 \\ &= X_2\delta + v_1 - V_2\delta \\ &= X_2\delta + \varepsilon_1. \end{aligned} \tag{23}$$

Therefore, under $H_0 : \mathrm{r}(\Pi) = k_x - 1$, we have that

$$E(z_i \varepsilon_{1i}) = E(z_i(v_{1i} - v_{2i}'\delta)) = 0, \tag{24}$$

as $E(z_i v_i') = 0$ from standard the linear projection results.

Let $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$, with Z_1 an $n \times (k_x - 1)$ matrix and Z_2 an $n \times (k_z - k_x + 1)$ matrix. Partition π_1 and Π_2 accordingly as $\pi_1 = \begin{bmatrix} \pi_{11}' & \pi_{12}' \end{bmatrix}'$ and $\Pi_2 = \begin{bmatrix} \Pi_{21}' & \Pi_{22}' \end{bmatrix}'$. Assume that Π_{21} has full rank $k_x - 1$, then, as in (8), we have

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 \\ &= X_2\Pi_{21}^{-1}\pi_{11} + Z_2\left(\pi_{12} - \Pi_{22}\Pi_{21}^{-1}\pi_{11}\right) + v_1 - V_2\Pi_{21}^{-1}\pi_{11} \\ &= X_2\kappa + Z_2\gamma + w_1, \end{aligned} \tag{25}$$

where $\kappa = \Pi_{21}^{-1}\pi_{11}$. If $\pi_1 = \Pi_2\delta$, then $\kappa = \delta$ and $\gamma = 0$. When $\pi_1 \neq \Pi_2\delta$ the overidentifying instruments are in general not "valid" instruments in (23) in the sense that $E(z_i\varepsilon_{1i}) \neq 0$. Note that the standard overidentification test is a score test for testing $H_0 : \gamma = 0$, as in (3).

The intuition of orthogonality condition (24) is clear. If the instruments are not correlated with ε_1, then they have no explanatory power to predict x_1 after having controlled for the other endogenous explanatory variables in the model.

Let $\widehat{\delta}_L$ and $\widehat{\Pi}_{2L}$ be the LIML estimators of δ and Π_2 in model (23). Let $\widehat{\varepsilon}_{1L} = x_1 - X_2\widehat{\delta}_L$ and $\widehat{X}_{2L} = Z\widehat{\Pi}_{2L}$.From the results derived above for the tests of overidentifying restrictions it follows that the non-robust Cragg-Donald and robust Kleibergen-Paap rank tests for $H_0 : \text{r}(\Pi) = k_x - 1$ can be obtained as (robust) tests for $H_0 : \gamma = 0$ in the specification

$$\widehat{\varepsilon}_{1L} = \widehat{X}_{2L}\eta + Z_2\gamma + \zeta_{1L}$$

after estimation by OLS.

Specifying the variance estimator of $\widehat{\gamma}$ only valid under conditional homoskedasticity and either based on the residual $\widehat{\zeta}_{1L} = \widehat{v}_1 - \widehat{V}_2\widehat{\delta}_L$ or $\widehat{\varepsilon}_{1L}$ results in the Basmann or Sargan version of the non-robust CD test

$$\begin{aligned} B\left(\widehat{\delta}_L\right) &= \frac{\widehat{\varepsilon}'_{1L}P_Z\widehat{\varepsilon}_{1L}}{\widehat{\zeta}'_{1L}\widehat{\zeta}_{1L}/n}; \\ S\left(\widehat{\delta}_L\right) &= \frac{\widehat{\varepsilon}'_{1L}P_Z\widehat{\varepsilon}_{1L}}{\widehat{\varepsilon}'_{1L}\widehat{\varepsilon}_{1L}/n}. \end{aligned}$$

Equivalently, specifying a robust variance estimator for $\widehat{\gamma}$ results in the robust versions of the Kleibergen-Paap test

$$B_r\left(\widehat{\delta}_L\right) = \widehat{\varepsilon}'_{1L}M_{\widehat{X}_{2L}}Z_2\left(Z'_2M_{\widehat{X}_{2L}}H_{\widehat{\zeta}_{1L}}M_{\widehat{X}_{2L}}Z_2\right)^{-1}Z'_2M_{\widehat{X}_{2L}}\widehat{\varepsilon}_{1L};$$

$$S_r\left(\widehat{\delta}_L\right) = \widehat{\varepsilon}'_{1L}M_{\widehat{X}_{2L}}Z_2\left(Z'_2M_{\widehat{X}_{2L}}H_{\widehat{\varepsilon}_{1L}}M_{\widehat{X}_{2L}}Z_2\right)^{-1}Z'_2M_{\widehat{X}_{2L}}\widehat{\varepsilon}_{1L}.$$

The robust versions of the CD test are obtained as the Hansen J-test in model (23) after CU-GMM estimation, as either the Basmann version $J_v\left(\widehat{\delta}_{cu,v}\right)$ or the Sargan version $J\left(\widehat{\delta}_{cu}\right)$. The test statistics have a limiting $\chi^2_{k_z - k_{x+1}}$ distribution under the null and maintained assumptions. An alternative to the robust KP and CD statistics is $J\left(\widehat{\delta}_{2L}\right)$, based on the two step LIML based procedure outlined above in (21), or $J\left(\widehat{\delta}_{2L,r}\right)$

There is a direct correspondence between $S\left(\widehat{\delta}_L\right)$ and the canonical correlation test of Anderson (1951). Denote the signed canonical correlations of x_i and z_i by ρ_j, $j = 1, ..., k_x$, ordered such that $\rho_1 \geq \rho_2 \geq ... \geq \rho_{k_x} \geq 0$. The canonical correlation test for $H_0 : \text{r}(\Pi) = k_x - 1$ is a test for $H_0 : \rho_{k_x} = 0$, see Anderson (1951). Let $\widehat{\rho}^2_{k_x}$ be the smallest sample squared canonical correlation, which is given by

$$\widehat{\rho}^2_{k_x} = \min \text{eval}\left((X'X)^{-1}X'P_ZX\right).$$

It follows that

$$S\left(\widehat{\delta}_L\right) = n\widehat{\rho}^2_{k_x} \xrightarrow{d} \chi^2_{k_z-k_x+1}, \tag{26}$$

under the null and maintained assumptions. The relationship between $B\left(\widehat{\delta}_L\right)$ and $S\left(\widehat{\delta}_L\right)$ is given by

$$B\left(\widehat{\delta}_L\right) = \frac{S\left(\widehat{\delta}_L\right)}{1-\widehat{\rho}^2_{k_x}},$$

from which it follows that $B\left(\widehat{\delta}_L\right) \geq S\left(\widehat{\delta}_L\right)$, with the discrepancy increasing with increasing value of $\widehat{\rho}^2_{k_x}$.

All the above tests are invariant to normalisation, i.e. the results are the same if an explanatory variable different from x_1 is chosen as the dependent variable in (23). In contrast, Sanderson and Windmeijer (2016) (SW) proposed conditional F-statistics for testing for underidentification or weak instruments for each endogenous variable separately. Their conditional tests statistics are based on the Basmann version of the underidentification test. Let $\widehat{\delta}_j$ be the 2SLS estimator of δ_j in the model

$$x_j = X_{-j}\delta_j + \varepsilon_j \tag{27}$$

for $j = 1, ..., k_x$, using instruments Z, where X_{-j} is X without x_j. Let $\widehat{\varepsilon}_j = x_j - X_{-j}\widehat{\delta}_j$. SW proposed use of the Basmann tests

$$B\left(\widehat{\delta}_j\right) = \frac{\widehat{\varepsilon}_j' P_Z \widehat{\varepsilon}_j}{\widehat{\zeta}_j'\widehat{\zeta}_j/n},$$

where $\widehat{\zeta}_j = \widehat{v}_j - \widehat{V}_{-j}\widehat{\delta}_j$, and provided the theory for testing for weak instruments based on the F-test version, $F_j = B\left(\widehat{\delta}_j\right)/(k_z - k_x + 1)$. The weak instrument asymptotics they considered was that of $\mathrm{r}(\Pi)$ local to a rank reduction of 1, or $\pi_j = \Pi_{-j}\delta_j + l/\sqrt{n}$. Clearly, instead of the Basmann tests, one could equivalently consider the Sargan versions $S\left(\widehat{\delta}_j\right)$.

SW showed that the conditional F_j can provide additional information to that provided by the CD statistic about the nature of the weak instruments problem. This is related to the cardinality, i.e. the number of non-zero elements of $\widehat{\delta}_*$. If $|\delta^*| = k_x$ then the null holds for all j in (27) and so none of the $B\left(\widehat{\delta}_j\right)$ will reject the null in large samples. If $|\delta^*| = k_x - s$, then s of the $B\left(\widehat{\delta}_j\right)$ will reject the null in large samples.

A simple generalisation to robust tests for underidentification is then to compute the robust versions $B_r\left(\widehat{\delta}_j\right)$ or $S_r\left(\widehat{\delta}_j\right)$. From the results in Section 3 if follows that the latter are simply the two-step Hansen J-tests in (27), for $j = 1, ..., k_x$.

5 Testing for Underidentification in Dynamic Panel Data Models

We consider an i.i.d. sample $\{y_i, X_i\}_{i=1}^n$, where y_i is the T-vector (y_{it}) and X_i is the $T \times k_x$ matrix $[x'_{it}]$. The linear panel data model is specified as

$$y_{it} = x'_{it}\beta + \eta_i + u_{it}$$

for $i = 1, ..., n$, $t = 1, ...T$, where x_{it} can contain lags of the dependent variable. The Arellano and Bond (1991) procedure to estimate the parameters β is to first-difference the model

$$\Delta y_{it} = (\Delta x_{it})' \beta + \Delta u_{it}$$

and estimate by GMM, using lagged levels of the explanatory and dependent variables as sequential instruments. Assuming that x_{it} contains a the lagged dependent variable and that all other explanatory variables are endogenous, the available moment conditions at period t are given by

$$E\left(x_i^{t-2}\Delta u_{it}\right), \tag{28}$$

where $x_i^{t-2} = \begin{pmatrix} x'_{i1} & x'_{i2} & \ldots & x'_{i,t-2} \end{pmatrix}'$. The moments (28) can be expressed as

$$E\left(Z'_i \Delta u_i\right) = 0,$$

for $i = 1, ..., n$, where Δu_i is the $(T-2)$-vector $(\Delta u_{it})_{t=3}^T$, and Z_i is the $(T-2) \times k_z$ matrix

$$Z_i = \begin{bmatrix} x_i^{1\prime} & 0 & \cdots & 0 \\ 0 & x_i^{2\prime} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & x_i^{T-2\prime} \end{bmatrix},$$

with $k_z = k_x (T-1)(T-2)/2$.

For testing underidentification in this setup, consider the first-stage linear projection model

$$\Delta X_i = Z_i \Pi + V_i,$$

where $\Delta X_i = \left[(x_{it} - x_{i,t-1})'\right]_{t=3}^{T}$ is a $(T-2)\times k_x$ matrix, Π is a $k_z \times k_x$ matrix and V_i is a $(T-2)\times k_x$ matrix. For the errors V_i we now have $E\left(\text{vec}\left(V_i\right)\text{vec}\left(V_i\right)'\right) = \Sigma_{\text{vec}(V)}$. Whilst we can still make an assumption of conditional homoskedasticity, $E\left(\text{vec}\left(V_i\right)\text{vec}\left(V_i\right)'|Z_i\right) = \Sigma_{\text{vec}(V)}$, it seems implausible to assume $\Sigma_{\text{vec}(V)} = \Sigma_v \otimes I_n$. For example, due to the nature of the sequential moments, the variances $E\left(v_{ijt}^2\right)$ will be varying over time.

Therefore, the non-robust version of the CD test for testing $H_0 : \text{r}\left(\Pi\right) = k_x - 1$ will be the minimum distance criterion based on a variance estimator of the OLS estimator $\widehat{\pi} = \text{vec}\left(\widehat{\Pi}\right)$, that takes the clustering into account whilst making an assumption of conditional homoskedasticity. As this variance does not have a kronecker representation, this no longer is a simple minimum eigenvalue problem and the solution needs to be obtained via iterative methods.

Partition $\Delta X_i = \left[\begin{array}{cc} \Delta x_{1i} & \Delta X_{2i} \end{array}\right]$ and $\Pi = \left[\begin{array}{cc} \pi_1 & \Pi_2 \end{array}\right]$. The robust CD statistic is obtained as the CU-GMM J-test statistic in the model

$$\Delta x_{1i} = \left(\Delta X_{2i}\right)\delta + \varepsilon_{1i}, \tag{29}$$

using instruments Z_i. The CU-GMM criterion is given by

$$J\left(\widehat{\delta}_{cu}\right) = \min_{\delta}\left(\sum_{i=1}^{n} Z_i'\varepsilon_{1i}\left(\delta\right)\right)'\left(\sum_{i=1}^{n} Z_i'\varepsilon_{1i}\left(\delta\right)\varepsilon_{1i}\left(\delta\right)'Z_i\right)^{-1}\left(\sum_{i=1}^{n} Z_i'\varepsilon_{1i}\left(\delta\right)\right),$$

where $\varepsilon_{1i}\left(\delta\right) = \Delta x_{1i} - \left(\Delta X_{2i}\right)\delta$. The CU-GMM estimator can be obtained from the iterative procedure described in Section 3.4, using a cluster robust variance estimator of $\widehat{\pi}$.

The LIML normalised robust Kleibergen-Paap test is based on the pooled LIML estimator of δ in (29), i.e. the estimator that would be efficient if $\Sigma_{\text{vec}(V)} = \Sigma_v \otimes I_n$. Let $\widehat{\delta}_L$ and $\widehat{\Pi}_{2L}$ again denote the LIML estimators of δ and Π_2, and let $\widehat{\varepsilon}_{1L,i} = \Delta x_{1i} - \Delta X_{2i}\widehat{\delta}_L$ and $\Delta\widehat{X}_{2L,i} = Z_i\widehat{\Pi}_{2L}$. Z_{2i} is any $k_z - k_x + 1$ subset of instruments. Then the cluster robust KP test is obtained as the test for $H_0 : \gamma = 0$ in the specification

$$\widehat{\varepsilon}_{1L,i} = \Delta\widehat{X}_{2L,i}\eta + Z_{2i}\gamma + \zeta_{1L,i},$$

estimated by OLS and specifying a cluster robust variance estimator for $\widehat{\gamma}$. Under the null that $\text{r}\left(\Pi\right) = k_x - 1$, these test statistics are invariant and have a limiting $\chi^2_{k_z - k_x + 1}$ distribution under the maintained assumptions. Although the pooled LIML estimator will not be an efficient estimator, this does not affect the local asymptotic power of the test.

The robust Sargan versions of the Sanderson-Windmeijer individual conditional underidentification tests is to estimate the specifications

$$\Delta x_{j,i} = (\Delta X_{-j,i})\, \delta_j + \varepsilon_{j,i}$$

by two-step GMM, again using Z_i as the instruments, and to test the null $H_0 : E(Z_i' \varepsilon_{j,i}) = 0$ with the Hansen J-test. These tests are particularly easy to perform with gmm estimation routines like xtabond2, as one simply has to perform the same estimation as for the original dependent variable, replacing the latter by one of the explanatory variables, and keeping the instrument specification the same. As an example, consider the model of interest that has as endogenous explanatory variables the lagged dependent variable and two further variables x_1 and x_2. The estimation command for the model using xtabond2 (Roodman, 2009) in Stata is

xtabond2 y l.y x1 x2 i.year, gmm(y x1 x2, lag (2 .)) iv(i.year) nol rob.

Then the per endogenous explanatory variable conditional robust underidentification tests can be obtained as the Hansen tests from the GMM estimation sequence

xtabond2 l.y x1 x2 i.year, gmm(y x1 x2, lag(2 .)) iv(i.year) nol rob;
xtabond2 x1 l.y x2 i.year, gmm(y x1 x2, lag(2 .)) iv(i.year) nol rob;
xtabond2 x2 l.y x1 i.year, gmm(y x1 x2, lag(2 .)) iv(i.year) nol rob.

The above methods extend straightforwardly to the system estimator of Blundell and Bond (1998).

Table 1 presents results for the underidentification tests for the production function estimation example of Blundell and Bond (2000). The estimated model is

$$y_{it} = \rho y_{i,t-1} + \beta_n n_{it} + \gamma_n n_{i,t-1} + \beta_k k_{it} + \gamma_k k_{i,t-1} + \delta_t + \eta_i + u_{it}, \tag{30}$$

where y_{it} is log sales of firm i in year t, n_{it} is log employment and k_{it} is log capital stock. The data used is a balanced panel of 509 R&D-performing US manufacturing companies observed for 8 years, 1982-89. Model specification (30) is a Cochran-Orcutt transformed model to deal with serial correlation of the errors in the static Cobb-Douglas production function. Although there is therefore more information due to the non-linear relationships between the parameters, $\gamma_n = -\rho\beta_n$ and $\gamma_k = -\rho\beta_k$, we ignore this information for illustrative purposes. The instruments specified by Blundell and Bond (2000) for the first-differenced model are the lagged levels of y, n and k, dated $t-3$ up till $t-5$. The $t-2$ lag was not used as an instrument due to measurement error problems in the variables. The System estimator then additionally includes $\Delta y_{i,t-2}$, $\Delta n_{i,t-2}$ and $\Delta k_{i,t-2}$ as per period instruments for the model in levels.

Table 1. P-values for underidentification tests

	$S_r\left(\widehat{\delta}_L\right)$	$S_r\left(\widehat{\delta}_j\right)$				
		$y_{i,t-1}$	n_{it}	$n_{i,t-1}$	k_{it}	$k_{i,t-1}$
1st Differences	0.527	0.001	0.153	0.008	0.108	0.018
System	0.178	0.002	0.119	0.041	0.000	0.001

The Kleibergen-Paap LIML based cluster robust score test $S_r\left(\widehat{\delta}_L\right)$ clearly indicates that the instruments do not identify the parameters in the first-differenced model with the p-value of the test equal to 0.527. From the individual Hansen J-test statistics $S_r\left(\widehat{\delta}_j\right)$, it becomes clear that the lagged levels instruments do not have predictive power for the endogenous variables Δn_{it} and Δk_{it}, after having to predict $\Delta n_{i,t-1}$, $\Delta k_{i,t-1}$ and $\Delta y_{i,t-1}$. The underidentification test for the System estimator also does not reject the null of underidentification, with a p-value of 0.178. The individual test statistics indicate that the instruments for the System estimator mainly fail to predict current log employment.

6 Testing for General Rank

We return to the original models (1), (2) and data $w_i = \left(\begin{array}{cc} y_i & x_i' \end{array}\right)'$ and consider testing a general null hypothesis on the rank of Π^*. For ease of exposition, we consider the situation where there are two linear relationships between the variables such that

$$E\left(z_i w_i' \left(\begin{array}{cc} \psi_1 & \psi_2 \end{array}\right)\right) = E\left(z_i w_i' \Psi\right) = 0$$

or $r\left(\Pi^*\right) = k_x - 1$. As in Arellano, Hansen and Sentana (2012), we start by standardising $\Psi = \left[\begin{array}{c} I_2 \\ B \end{array}\right]$. Partition $w = \left[\begin{array}{ccc} y & x_1 & X_2 \end{array}\right]$ and $\Pi = \left[\begin{array}{ccc} \pi_y & \pi_1 & \Pi_2 \end{array}\right]$. We then have the two equations

$$y = X_2\beta_y + u_y \tag{31}$$

$$x_1 = X_2\beta_x + u_x \tag{32}$$

and the test for overidentifying restrictions is a test for $H_0 : E(z_i u_i') = 0$, where here $u_i = \begin{pmatrix} u_{y,i} & u_{x,i} \end{pmatrix}'$.

Let Z_2 be any selection of $k_z - k_x + 1$ instruments, and specify

$$\begin{aligned} y &= X_2\beta_y + Z_2\gamma_y + u_y \\ x_1 &= X_2\beta_x + Z_2\gamma_x + u_x, \end{aligned}$$

then a test for $H_0 : E(z_i u_i') = 0$ is a score test for $H_0 : \gamma_y = \gamma_x = 0$. Both equations are again just identified, and hence the IV estimators for γ_y and γ_x are given by $\begin{pmatrix} \widehat{\gamma}_y & \widehat{\gamma}_x \end{pmatrix} = \left(Z_2' M_{\widehat{X}_2} Z_2\right)^{-1} Z_2' M_{\widehat{X}_2} \begin{pmatrix} y & x_1 \end{pmatrix}$, where $\widehat{X}_2 = Z\widehat{\Pi}_2$. These IV estimators are also efficient under conditional homoskedasticity, $Var(u_i|z_i) = \Sigma_u$, by standard SURE arguments.

Let $\widehat{\beta}_1 = \begin{pmatrix} \widehat{\beta}_{y,1}' & \widehat{\beta}_{x,1}' \end{pmatrix}'$, with $\widehat{\beta}_{y,1}$ and $\widehat{\beta}_{x,1}$ initial IV/GMM estimators of β_y and β_x in the restricted models (31) and (32), with $\widehat{u}_1 = \begin{pmatrix} \widehat{u}_{y,1}' & \widehat{u}_{x,1}' \end{pmatrix}'$ the associated residuals. Analogous to the test derived in Section 3, the robust score test for $H_0 : \gamma_y = \gamma_x = 0$ is then given by

$$\begin{aligned} S_r\left(\widehat{\beta}_1\right) &= \begin{pmatrix} \widehat{\gamma}_y \\ \widehat{\gamma}_x \end{pmatrix}' \left(V\widehat{a}r_{r,\widehat{u}_1}\begin{pmatrix} \widehat{\gamma}_y \\ \widehat{\gamma}_x \end{pmatrix}\right)^{-1} \begin{pmatrix} \widehat{\gamma}_y \\ \widehat{\gamma}_x \end{pmatrix} \\ &= \begin{pmatrix} y \\ x_1 \end{pmatrix}' \widetilde{Z}_2 \left(\widetilde{Z}_2' H_{\widehat{u}_1} \widetilde{Z}_2\right)^{-1} \widetilde{Z}_2' \begin{pmatrix} y \\ x_1 \end{pmatrix} \\ &= \begin{pmatrix} \widehat{u}_{y,1} \\ \widehat{u}_{x_1,1} \end{pmatrix}' \widetilde{Z}_2 \left(\widetilde{Z}_2' H_{\widehat{u}_1} \widetilde{Z}_2\right)^{-1} \widetilde{Z}_2' \begin{pmatrix} \widehat{u}_{y,1} \\ \widehat{u}_{x_1,1} \end{pmatrix}, \end{aligned} \tag{33}$$

where $\widetilde{Z}_2 = I_2 \otimes M_{\widehat{X}_2} Z_2$, $\frac{1}{\sqrt{n}}\widetilde{Z}_2' u \rightarrow N(0, \Omega_{\widetilde{z}_2 u})$, where $u = \left(u_y', u_x'\right)'$ and $n^{-1}\left(\widetilde{Z}_2' H_{\widehat{u}_1} \widetilde{Z}_2\right)$ is a consistent estimator of $\Omega_{\widetilde{z}_2 u}$. Under the maintained assumptions and H_0, $S_r\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{2(k_z-k_x+1)}$.

The 2SLS based non-robust version which is valid in the homoskedastic case, has $\widetilde{Z}_2' H_{\widehat{u}_{2sls}} \widetilde{Z}_2 = \widehat{\Sigma}_{\widehat{u}_{2sls}} \otimes Z_2' M_{\widehat{X}_2} Z_2$, and so

$$\begin{aligned} S\left(\widehat{\beta}_{2sls}\right) &= \widehat{u}_{2sls}' \left(\widehat{\Sigma}^{-1}_{\widehat{u}_{2sls}} \otimes M_{\widehat{X}_2} Z_2 \left(Z_2' M_{\widehat{X}_2} Z_2\right)^{-1} Z_2' M_{\widehat{X}_2}\right) \widehat{u}_{2sls} \\ &= \widehat{u}_{2sls}' \left(\widehat{\Sigma}^{-1}_{\widehat{u}_{2sls}} \otimes P_Z\right) \widehat{u}_{2sls}. \end{aligned}$$

Let $\dot{Z} = (I_2 \otimes Z)$. For a general one-step estimator $\widehat{\beta}_1$, the two-step GMM estimator is given by

$$\widehat{\beta}_2 = \arg\min_{\beta_y,\beta_x} \begin{pmatrix} y - X_2\beta_y \\ x_1 - X_2\beta_x \end{pmatrix}' \dot{Z}\left(\dot{Z}'H_{\widehat{u}_1}\dot{Z}\right)^{-1}\dot{Z}' \begin{pmatrix} y - X_2\beta_y \\ x_1 - X_2\beta_x \end{pmatrix},$$

with the Hansen J-test given by

$$\begin{aligned} J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) &= \widehat{u}_2'\dot{Z}\left(\dot{Z}'H_{\widehat{u}_1}\dot{Z}\right)^{-1}\dot{Z}'\widehat{u}_2 \\ &= S_r\left(\widehat{\beta}_1\right). \end{aligned}$$

Next let $\widehat{\psi}_{1L}$ and $\widehat{\psi}_{2L}$ be the LIML estimates of ψ_1 and ψ_2. These are obtained as $\widehat{\psi}_{1L} = \widehat{\Sigma}_w^{-1/2}v_{[1]}$ and $\widehat{\psi}_{2L} = \widehat{\Sigma}_w^{-1/2}v_{[2]}$ where $v_{[1]}$ and $v_{[2]}$ are the orthonormal eigenvectors associated with the 2 smallest eigenvalues of $\widehat{\Sigma}_w^{-1/2}W'P_ZW\widehat{\Sigma}_w^{-1/2}$. These estimates therefore have the normalisation $\widehat{\psi}_{1L}'\widehat{\Sigma}_w\widehat{\psi}_{1L} = \widehat{\psi}_{2L}'\widehat{\Sigma}_w\widehat{\psi}_{2L} = 1$ and $\widehat{\psi}_{1L}'\widehat{\Sigma}_w\widehat{\psi}_{2L} = 0$. Let $\widehat{u}_L = \text{vec}\left(W\widehat{\Psi}_L\right)$, where $\widehat{\Psi}_L = \begin{bmatrix} \widehat{\psi}_{1L} & \widehat{\psi}_{2L} \end{bmatrix}$. Then the Sargan, non-robust version of the score test is given by

$$S\left(\widehat{\Psi}_L\right) = \widehat{u}_L'\left(\widehat{\Sigma}_{\widehat{u}_L}^{-1} \otimes P_Z\right)\widehat{u}_L.$$

However, as

$$\widehat{\Sigma}_{\widehat{u}_L} = \frac{1}{n}\widehat{\Psi}_L'W'W\widehat{\Psi}_L = I_2,$$

it follows that

$$S\left(\widehat{\Psi}_L\right) = e_{[1]} + e_{[2]},$$

the sum of the two smallest eigenvalues of $\widehat{\Sigma}_w^{-1/2}W'P_ZW\widehat{\Sigma}_w^{-1/2}$.

Next, partition $\widehat{\Psi}_L = \begin{bmatrix} \widehat{\Psi}_{AL} \\ \widehat{\Psi}_{BL} \end{bmatrix}$ where $\widehat{\Psi}_{AL}$ is a 2×2 matrix, and let $\widehat{\Psi}_L^* = \widehat{\Psi}_L\widehat{\Psi}_{AL}^{-1} = \begin{bmatrix} I_2 \\ \widehat{\Psi}_{BL}\widehat{\Psi}_{AL}^{-1} \end{bmatrix}$. Then $\widehat{u}_L^* = \text{vec}\left(W\widehat{\Psi}_L^*\right) = \left(\left(\widehat{\Psi}_{AL}^{-1}\right)' \otimes I_2\right)\widehat{u}_L$, and $\widehat{\Sigma}_{\widehat{u}_L^*} = \frac{1}{n}\widehat{\Psi}_L^{*\prime}W'W\widehat{\Psi}_L^* = \widehat{\Psi}_{AL}^{-1\prime}\widehat{\Psi}_{AL}^{-1}$, and so

$$\begin{aligned} S\left(\widehat{\Psi}_L^*\right) &= \widehat{u}_L^{*\prime}\left(\widehat{\Sigma}_{\widehat{u}_L^*}^{-1} \otimes P_Z\right)\widehat{u}_L^* \\ &= \widehat{u}_L'\left(\left(\widehat{\Psi}_{AL}^{-1}\right) \otimes I_n\right)\left(\left(\widehat{\Psi}_{AL}\widehat{\Psi}_{AL}'\right) \otimes P_Z\right)\left(\left(\widehat{\Psi}_{AL}^{-1}\right)' \otimes I_n\right)\widehat{u}_L \\ &= S\left(\widehat{\Psi}_L\right). \end{aligned}$$

We can now link this to the Cragg-Donald minimum distance criterion, with the result that

$$S\left(\widehat{\Psi}_L^*\right) = \min_{\Pi_2,\beta_y,\beta_x} \begin{pmatrix} \widehat{\pi}_y - \Pi_2\beta_y \\ \widehat{\pi}_1 - \Pi_2\beta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix}' \left(\widehat{\Sigma}_w^{-1} \otimes (Z'Z)\right) \begin{pmatrix} \widehat{\pi}_y - \Pi_2\beta_y \\ \widehat{\pi}_1 - \Pi_2\beta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix},$$

and the resulting estimators $\left[\begin{array}{cc} \widehat{\beta}_{yL} & \widehat{\beta}_{xL} \end{array}\right] = \widehat{\Psi}_{BL}\widehat{\Psi}_{AL}^{-1}$, see also Cragg and Donald (1993). Clearly, this is the invariant LIML based rank test for $H_0 : \text{r}(\Pi^*) = k_x - 1$ against $H_1 : \text{r}(\Pi^*) > k_x - 1$.

Let $\widehat{\Pi}_{2L}$ be the LIML estimator of Π_2, and let $\widehat{X}_{2L} = Z\widehat{\Pi}_{2L}$ The LIML based Kleibergen-Paap rank test is then the robust score test

$$\begin{aligned} S_r\left(\widehat{\beta}_L\right) &= \widehat{u}_L^{*\prime}\widetilde{Z}_{2L}\left(\widetilde{Z}_{2L}'H_{\widehat{u}_L^*}\widetilde{Z}_{2L}\right)^{-1}\widetilde{Z}_{2L}'\widehat{u}_L^* \\ &= \widehat{u}_L'\widetilde{Z}_{2L}\left(\widetilde{Z}_{2L}'H_{\widehat{u}_L}\widetilde{Z}_{2L}\right)^{-1}\widetilde{Z}_{2L}'\widehat{u}_L. \end{aligned}$$

where $\widetilde{Z}_{2L} = \left(I_2 \otimes M_{\widehat{X}_{2L}} Z_2\right)$. Note that, as before, the estimator $\widehat{\Pi}_{2L}$ can be obtained directly from the minimum eigenvalue LIML estimator $\widehat{\Psi}_L$. Let $\widehat{U}_L = W\widehat{\Psi}_L$, then

$$\widehat{\Pi}_{2L} = \left(Z'M_{\widehat{U}_L}Z\right)^{-1} Z'M_{\widehat{U}_L}X_2.$$

The CU-GMM robust invariant CD rank test is

$$J\left(\widehat{\beta}_{cu}\right) = S_r\left(\widehat{\beta}_{cu}\right) = \min_{\Pi_2,\beta_y,\beta_x} \begin{pmatrix} \widehat{\pi}_y - \Pi_2\beta_y \\ \widehat{\pi}_1 - \Pi_2\beta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix}' \left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi_2\beta_y \\ \widehat{\pi}_1 - \Pi_2\beta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix}.$$

These tests are versions of the Arellano, Hansen and Sentana (2012) I test for underidentification for the standard linear IV model. They are easily generalised to testing for general $H_0 : \text{r}(\Pi^*) = q$ against $H_1 : \text{r}(\Pi^*) > q$, which are score type tests of the form (33), with the LIML and CU-GMM versions being invariant to normalisation.

It is at this point illustrative to consider Example 3.4 in AHS (2012, pp. 262-263). They considered a normalized four-input translog cost share equation system, resulting in the equations

$$y_{j,t} = \beta_{j,1}p_{1,t} + \beta_{j,2}p_{2,t} + \beta_{j,3}p_{3,t} + v_{j,t}$$

for $j = 1, 2, 3$, $t = 1, ..., T$, and where $y_{j,t}$ denotes the cost share of input j and $p_{j,t}$ is the log price of input j relative to the omitted input, and $w_t = (y_{1,t}\ y_{2,t}\ y_{3,t}\ p_{1,t}\ p_{2,t}\ p_{3,t})'$. The symmetry constraints are given by

$$\beta_{j,k} = \beta_{k,j} \quad j \neq k.$$

Prices are endogenous and there is a k_z dimensional vector of instruments z_t available to instrument prices under the assumption that $E(z_t v_{j,t}) = 0$ for $j = 1, 2, 3$. For this case we have $\Pi^* = \begin{bmatrix} \pi_{y_1} & \pi_{y_2} & \pi_{y_3} & \pi_{p_1} & \pi_{p_2} & \pi_{p_3} \end{bmatrix} = \begin{bmatrix} \Pi_y & \Pi_p \end{bmatrix}$. The test for overidentifying restrictions $H_0 : E(z_t v_{j,t}) = 0$ for $j = 1, 2, 3$ is in this case a test for $H_0 : \mathrm{r}(\Pi^*) = 3$, which, incorporating the restrictions, can for example be obtained as the CU-GMM criterion

$$J\left(\widehat{B}_{cu}\right) = \min_{\Pi_p, B} \begin{pmatrix} \mathrm{vec}\left(\widehat{\Pi}_y - \Pi_p B\right) \\ \mathrm{vec}\left(\widehat{\Pi}_p - \Pi_p\right) \end{pmatrix}' \left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \mathrm{vec}\left(\widehat{\Pi}_y - \Pi_p B\right) \\ \mathrm{vec}\left(\widehat{\Pi}_p - \Pi_p\right) \end{pmatrix},$$

with

$$B = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{1,2} & \beta_{2,2} & \beta_{2,3} \\ \beta_{1,3} & \beta_{2,3} & \beta_{3,3} \end{bmatrix}.$$

As Π^* is a $k_z \times 6$ matrix, it is clear that the necessary order condition is that $k_z \geq 3$. Note that the degrees of freedom of the test is equal to $k_z \times (6 - 3) - (9 - 3)$, so even if $k_z = 3$, the model is overidentified due to the symmetry restrictions.

Next, consider the AHS underidentification test $H_0 : \mathrm{r}(\Pi^*) = 2$. Let $\Pi_{yp_1} = \begin{bmatrix} \Pi_y & \pi_{p_1} \end{bmatrix}$ and $\Pi_{p2} = \begin{bmatrix} \pi_{p_2} & \pi_{p_3} \end{bmatrix}$. We now add a linear relationship of the form

$$p_{1,t} = \delta_2 p_{2,t} + \delta_3 p_{3,t} + \varepsilon_{1,t}$$

and express the original equations in terms of $p_{2,t}$ and $p_{3,t}$ only. Hence,

$$y_{j,t} = \left(\beta_{j,2} + \delta_2 \beta_{j,1}\right) p_{2,t} + \left(\beta_{j,3} + \delta_3 \beta_{j,1}\right) p_{3,t} + v_{j,t} + \beta_{j,1} \varepsilon_{1,t}.$$

Then the robust rank test for $H_0 : \mathrm{r}(\Pi^*) = 2$, incorporating all restrictions, is given by

$$J\left(\widehat{E}_{cu}\right) = \min_{\Pi_{p2}, E} \begin{pmatrix} \mathrm{vec}\left(\widehat{\Pi}_{yp_1} - \Pi_{p2} E\right) \\ \mathrm{vec}\left(\widehat{\Pi}_{p2} - \Pi_{p2}\right) \end{pmatrix}' \left(V\widehat{a}r_{r,w}\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \mathrm{vec}\left(\widehat{\Pi}_{yp_1} - \Pi_{p2} E\right) \\ \mathrm{vec}\left(\widehat{\Pi}_{p2} - \Pi_{p2}\right) \end{pmatrix},$$

with

$$\begin{aligned} E &= \begin{bmatrix} \beta_{1,2} + \delta_2 \beta_{1,1} & \beta_{2,2} + \delta_2 \beta_{1,2} & \beta_{2,3} + \delta_2 \beta_{1,3} & \delta_2 \\ \beta_{1,3} + \delta_3 \beta_{1,1} & \beta_{2,3} + \delta_3 \beta_{1,2} & \beta_{3,3} + \delta_3 \beta_{1,3} & \delta_3 \end{bmatrix} \\ &= \begin{bmatrix} e_{1,1} & e_{2,1} & e_{3,1} & e_{4,1} \\ e_{1,2} & e_{2,2} & e_{3,2} & e_{4,2} \end{bmatrix} \end{aligned}$$

and the restriction $(e_{3,1} - e_{2,2}) = e_{4,1}e_{1,2} - e_{4,2}e_{1,1}$.

This shows that the parameter restrictions in the linear model can be incorporated directly into the CD rank test procedure. Whilst the null hypothesis of the AHS test is clear, the main issue with this test seems to be that it is unclear what a rejection of the null implies. Rejecting the null $H_0 : \mathrm{r}(\Pi^*) = 2$ in the example above does not necessarily mean that the model is meaningfully identified, as it could well be the case that $E(z_t \varepsilon_{t,1}) = E(z_t v_{2,t}) = E(z_t v_{3,t}) = 0$, but $E(z_t v_{1,t}) \neq 0$. As what matters for identification in this model is whether the instruments can predict the endogenous prices, the more natural test for underidentification seems to be $H_0 : E(z_t \varepsilon_{1,t}) = 0$ or $H_0 : \mathrm{r}(\Pi_p) = 2$ against $H_1 : \mathrm{r}(\Pi_p) = 3$.

7 Testing the Rank of Parameter Matrices Estimated by OLS

The LIML and CU-GMM based rank tests on the matrices Π^* and Π may appear to be specific to the linear IV setup, as the instruments Z are used as instruments for both the over- and underidentification tests. This approach can, however, be applied to more general settings testing the rank on parameter matrices that are estimated by OLS. Consider for example the linear model specification as in Al-Sadoon (2017)

$$y_i = B' x_i + v_i$$

for $i = 1, ..., n$, where y_i and v_i are k_y-vectors, x_i is a k_x-vector and B is a $k_x \times k_y$ matrix of unknown parameters. In matrix notation the model is

$$Y = XB + V,$$

where Y is the $n \times k_y$ matrix $[y_i']$, X is the $n \times k_x$ matrix $[x_i']$ and V is the $n \times k_y$ matrix $[v_i']$. It is assumed that $E(x_i v_i') = 0$ and therefore the OLS estimator for B, $\widehat{B} = (X'X)^{-1} X'Y$ is consistent, with further regularity conditions in place for standard limiting normal distribution results.

Consider first the situation where $k_x \geq k_y$ and testing the null hypothesis $H_0 : \mathrm{r}(B) = k_y - 1$. This is the setup as in Cragg and Donald (1993, 1997), and in analogy to the IV results above, the CD rank test is a LIML/CU-GMM based score test. Partition

$Y = \left[\begin{array}{cc} y_1 & Y_2 \end{array} \right]$, $B = \left[\begin{array}{cc} b_1 & B_2 \end{array} \right]$, $\widehat{B} = \left[\begin{array}{cc} \widehat{b}_1 & \widehat{B}_2 \end{array} \right]$ and $X = \left[\begin{array}{cc} X_1 & X_2 \end{array} \right]$ where X_2 is any $(k_x - k_y + 1)$ subset of variables in X. Then the rank test is the score test for $H_0 : \gamma = 0$ in the specification

$$y_1 = Y_2 \delta + X_2 \gamma + \varepsilon_1.$$

Let $\widehat{\delta}_L$ and $\widehat{B}_{2L}$ be the LIML estimators of δ and B_2 in the restricted model

$$y_1 = Y_2 \delta + \varepsilon_1, \tag{34}$$

using X as instruments. These can be obtained from the minimum eigenvalue solution and projection as in Section 3.1, or as

$$\left(\widehat{\delta}_L, \widehat{B}_{2L} \right) = \arg \min_{\delta, B_2} \left(\begin{array}{c} \widehat{b}_1 - B_2 \delta \\ \operatorname{vec}\left(\widehat{B}_2 - B_2 \right) \end{array} \right)' \left(\widehat{\Sigma}_y^{-1} \otimes (X'X) \right) \left(\begin{array}{c} \widehat{b}_1 - B_2 \delta \\ \operatorname{vec}\left(\widehat{B}_2 - B_2 \right) \end{array} \right),$$

where $\widehat{\Sigma}_y = Y'Y/n$ and $\widehat{\Sigma}_y \otimes (X'X)^{-1}$ is the estimator of $Var\left(\operatorname{vec}\left(\widehat{B} \right) \right)$ under conditional homoskedasticity and $B = 0$.

Let $\widehat{Y}_{2L} = X \widehat{B}_{2L}$ and $\widehat{\varepsilon}_{1L} = y_1 - Y_2 \widehat{\delta}_L$, then the non-robust CD and robust KP tests can again be obtained from the tests for $H_0 : \gamma = 0$ in the specification

$$\widehat{\varepsilon}_{1L} = \widehat{Y}_{2L} \eta + X_2 \gamma + \zeta_{1L},$$

estimated by OLS. The robust CD test is $J\left(\widehat{\delta}_{cu} \right)$ after estimation of (34) by CU-GMM. The extensions to general rank tests, $H_0 : \mathrm{r}(B) = q$, are then as discussed in the previous section.

Next, consider the case where $k_x \leq k_y$. In that case, the CD and KP rank tests apply to the column rank of the $k_y \times k_x$ matrix B', for example $H_0 : \mathrm{r}(B') = k_x - 1$. The OLS estimator is then $\widehat{B}' = Y'X(X'X)^{-1}$ and the estimator of $Var\left(\operatorname{vec}\left(\widehat{B}' \right) \right)$ under conditional homoskedasticity and $B = 0$ is given by $(X'X)^{-1} \otimes \widehat{\Sigma}_y$. The KP LIML normalisation, $\widehat{\Theta}_{B'} = G \widehat{B}' F'$, is then obtained with $G'G = \widehat{\Sigma}_y^{-1}$ and $F'F = \widehat{\Sigma}_x = X'X/n$. Choosing wlog $G = G' = \widehat{\Sigma}_y^{-1/2}$ and $F = F' = \widehat{\Sigma}_x^{1/2}$ results in $\widehat{\Theta}_{B'} = (Y'Y)^{-1/2} Y'X (X'X)^{-1/2}$.

Next, consider the specification

$$X = YC + U,$$

with C a $k_y \times k_x$ matrix, the same dimension of B'. The OLS estimator is given by $\widehat{C} = (Y'Y)^{-1} Y'X$. Assuming conditional homoskedasticity and $C = 0$, the estimator for the variance of vec $\left(\widehat{C}\right)$ is given by $\widehat{\Sigma}_x \otimes (Y'Y)^{-1}$. For testing hypotheses on the rank of C, the KP LIML normalisation is then $\widehat{\Theta}_C = G\widehat{C}F'$, with here $G = \widehat{\Sigma}_y^{1/2}$ and $F = \widehat{\Sigma}_x^{-1/2}$. Hence $\widehat{\Theta}_C = (Y'Y)^{-1/2} Y'X (X'X)^{-1/2} = \widehat{\Theta}_{B'}$. Therefore, for this case where $k_y \leq k_x$, the CD and KP rank tests for, for example, $H_0 : \mathrm{r}(B') = k_x - 1$ are identical to the rank tests for $H_0 : \mathrm{r}(C) = k_x - 1$. Partition $X = \begin{bmatrix} x_1 & X_2 \end{bmatrix}$, then the tests can be obtained analogous to above by estimating the model

$$x_1 = X_2 \delta + \varepsilon_1$$

by LIML or CU-GMM, now using Y as the instruments. When $k_y = k_x$ the two approaches are identical.

This setup applies to the asset pricing model as further described below in Section 9. Kleibergen and Paap (2006) considered tests on the rank of the matrix Λ' in the specification

$$r_t = \Lambda' x_t + v_t,$$

for $t = 1, ..., T$, with r_t a k_r-vector of portfolio returns and $x_t = \begin{pmatrix} 1 & f_t' \end{pmatrix}'$, with f_t a k_f-vector of systematic risk factors, and so $k_x = k_f + 1$. In their setting $k_r > k_x$. Then it follows form the results above that the CD and KP rank tests for $H_0 : \mathrm{r}(\Lambda') = k_x - 1$ are the same as the rank test for $H_0 : r(C) = k_x - 1$ in the specification

$$x_t = C' r_t + u_t.$$

As $X = \begin{bmatrix} \iota_T & F \end{bmatrix}$, where ι_T is a T-vector of ones, and F is the $T \times k_f$ matrix of factors, these rank tests can be obtained as the score tests described above by estimating the specification

$$\iota_t = F\delta + \varepsilon$$

by LIML or CU-GMM, using here the $T \times k_r$ matrix of returns R as the matrix of instruments.

8 Limiting Distribution of the Sargan Test in Underidentified Models

The limiting distributions of the Sargan tests for overidentifying restrictions, $S\left(\widehat{\beta}_{2sls}\right)$ and $S\left(\widehat{\beta}_L\right)$, when $\mathrm{r}\left(z_i w_i'\right)$, or $\mathrm{r}\left(\Pi^*\right)$, is less then k_x have been derived by Kitamura (2005) for $S\left(\widehat{\beta}_{2sls}\right)$, and the result of Gospodinov, Kan and Robotti (2017) (GKR) derived in the context of linear factor models applies to $S\left(\widehat{\beta}_L\right)$. As $S\left(\widehat{\beta}_L\right)$ is an invariant rank test, its limiting distribution is determined by $\mathrm{r}\left(\Pi^*\right)$ only, independent of whether the moments restrictions $E\left(z_i u_i\right) = 0$ hold or not. In contrast, the limiting distribution of $S\left(\widehat{\beta}_{2sls}\right)$ under rank deficiency depends on whether the moment restrictions hold or not. These limiting distribution results hold under Assumptions 1, 2, 4 and 5 and the maintained assumption of conditional homoskedasticity. We focus here on the Sargan version of the tests, but the results equally apply to the Basmann versions.

Theorem 2 of GKR states the limiting distribution result for an asset-pricing model with linear moment restrictions. From the proof (GKR, p. 1626) it follows directly that the result holds for the minimum eigenvalue representation of $S\left(\widehat{\beta}_L\right)$ as given in (11). Let $\mathrm{r}\left(\Pi^*\right) = k_x + 1 - d$, for an integer d. Then the result is that for $d \geq 1$,

$$S\left(\widehat{\beta}_L\right) \xrightarrow{d} w_d, \tag{35}$$

where w_d is the smallest eigenvalue of $W_d \sim \mathrm{W}_d\left(k_z - k_x - 1 + d, I_d\right)$, and $\mathrm{W}_d\left(k_z - k_x - 1 + d, I_d\right)$ denotes the Wishart distribution with $k_z - k_x - 1 + d$ degrees of freedom and scaling matrix I_d.

When the moment conditions are valid, $E\left(z_i u_i\right) = 0$, the result for $S\left(\widehat{\beta}_{2sls}\right)$ as given in Theorem 3.1 in Kitamura (2005, p 67) is,

$$S\left(\widehat{\beta}_{2sls}\right) \xrightarrow{d} C \times B_d$$

where $C \sim \chi^2_{k_z - k_x}$, $B_d \sim \mathrm{Beta}\left(\frac{k_z - k_x + 1}{2}, \frac{d-1}{2}\right)$ and C and B_d are independent. As before, $\mathrm{r}\left(\Pi^*\right) = k_x + 1 - d$ with here $d \geq 2$. When $d = 1$, $B_d = 1$.

When the moment conditions are invalid, the result for $S\left(\widehat{\beta}_{2sls}\right)$ as given in Theorem 3.2 in Kitamura (2005, p 71) is,

$$S\left(\widehat{\beta}_{2sls}\right) \xrightarrow{d} C \times IB_d,$$

where $C \sim \chi^2_{k_z-k_x}$, $IB_d \sim$ Inverted Beta $\left(\frac{d}{2}, \frac{k_z-k_x+1}{2}\right)$ and C and IB_d are independent, with here $d \geq 1$.

Figure 1 displays the limiting distributions of $S\left(\widehat{\beta}_L\right)$ and $S\left(\widehat{\beta}_{2sls}\right)$ for $k_z - k_x = 7$, for $S\left(\widehat{\beta}_{2sls}\right)$ when the moment restrictions are valid, for values of $d = 1, 2, 3$. Figure 2 presents the limiting distribution of $S\left(\widehat{\beta}_{2sls}\right)$ when the moment conditions $E\left(z_i u_i\right) = 0$ are invalid, for the same values of d.

The densities for the LIML estimator are the same as in GKR, Figure 1, as the degrees of freedom are the same. Clearly, with rank deficiency, the rejection probability for both $S\left(\widehat{\beta}_L\right)$ and $S\left(\widehat{\beta}_{2sls}\right)$ is less than nominal size, with the discrepancy larger for $S\left(\widehat{\beta}_L\right)$ than for $S\left(\widehat{\beta}_{2sls}\right)$. Also, as $S\left(\widehat{\beta}_L\right)$ is an invariant rank test, it has power equal to size if the moment conditions $E\left(z_i u_i\right) = 0$ do not hold, but $\mathrm{r}\left(\Pi^*\right) = k_x$.

In contrast, Figure 2 shows that when the moment conditions do not hold, the limiting distribution of $S\left(\widehat{\beta}_{2sls}\right)$ under rank deficiency is very different. Although the test is no longer consistent, it still has power to reject the null. For this design, the power of the test at the 5% level in the limit is 0.881, 0.742 and 0.606 for $d = 1$, 2 and 3 respectively.

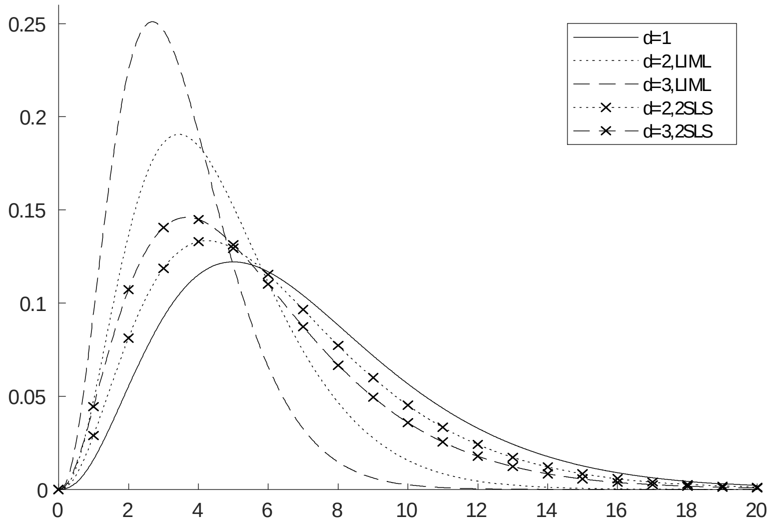

Figure 1. Limiting distributions of $S\left(\widehat{\beta}_L\right)$ and $S\left(\widehat{\beta}_{2sls}\right)$, for $\mathrm{r}\left(\Pi^*\right) = k_x + 1 - d$, $k_z - k_x = 7$. Moment conditions $E\left(z_i u_i\right) = 0$ are valid for $S\left(\widehat{\beta}_{2sls}\right)$.

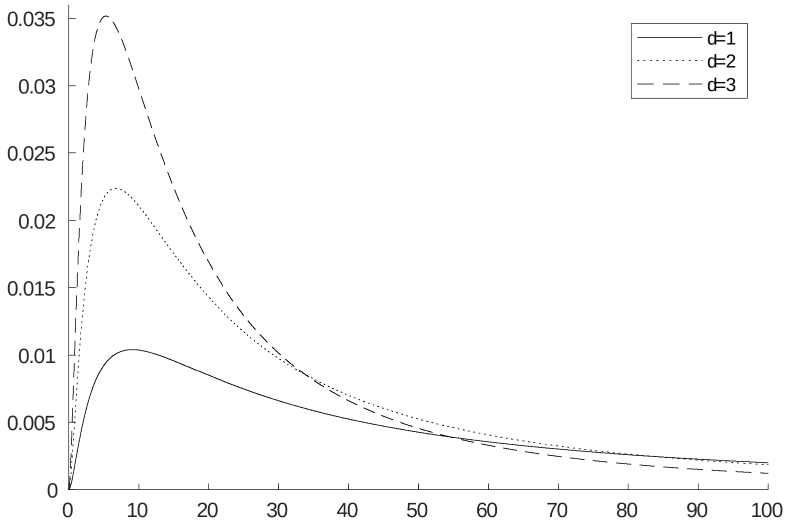

Figure 2. Limiting distributions of $S\left(\widehat{\beta}_{2sls}\right)$ for $\mathrm{r}\left(\Pi^{*}\right)=k_x+1-d$, $k_z-k_x=7$.
Moment conditions $E\left(z_i u_i\right)=0$ are invalid.

Table 2 presents the power of $S\left(\widehat{\beta}_{2sls}\right)$ to reject the null in the limit at the 5% level, for various combinations of the degrees of freedom $k_z-k_x=[1,..,5,10,15,20]$, and rank deficiency $d=[1,..,5]$. As Kitamura (2005, p. 74) shows, the limiting distribution of $S\left(\widehat{\beta}_{2sls}\right)$ when the moment conditions do not hold is equal to the distribution of $C_{k_z-k_x}C_{k_z-k_x+1}/C_d$, where the independent random variables are distributed as $C_{k_z-k_x}\sim\chi^2_{k_z-k_x}$, $C_{k_z-k_x+1}\sim\chi^2_{k_z-k_x+1}$ and $C_d\sim\chi^2_d$. We obtained the rejection probabilities from $1,000,000$ draws of the three random variables. As is clear from the results, the power of the test is increasing in k_z-k_x and decreasing in d. When $k_z-k_x=20$, the power is close to one, at 0.996, when $d=1$, and is still 0.917 when $d=5$. For the Monte Carlo analysis of GKR, as further detailed in Section 9, the degree of overidentification is around 20.

It is clear from the above results that the invariant $S\left(\widehat{\beta}_L\right)$ does not have power to reject $H_0: E\left(z_i u_i\right)=0$ in underidentified models when the moment conditions are invalid. This is clearly problematic when using $S\left(\widehat{\beta}_L\right)$ as a test for overidentifying restrictions. However, this is of course less of a problem for the use of $S\left(\widehat{\delta}_L\right)$ as a test for underidentification, and is also a reason why an underidentification test should be reported. The underidentification tests described in Section 4 are for testing $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) > k_x - 1$. If $\mathrm{r}(\Pi) < k_x - 1$, the limiting distribution results given in (35) apply to $S\left(\widehat{\delta}_L\right)$ and so the rejection frequency of $S\left(\widehat{\delta}_L\right)$ will be less than nominal size. This implies that a higher degree of underidentification does not lead to erroneous conclusions for the test of underidentification.

Table 2. Rejection probabilities, $P\left(S\left(\widehat{\beta}_{2sls}\right) > \chi^2_{k_z-k_x,0.95}\right)$

	d				
$k_z - k_x$	1	2	3	4	5
1	0.342	0.162	0.089	0.053	0.033
2	0.518	0.293	0.176	0.110	0.072
3	0.640	0.412	0.269	0.180	0.123
4	0.730	0.517	0.364	0.257	0.182
5	0.796	0.607	0.453	0.336	0.248
10	0.947	0.868	0.773	0.672	0.576
15	0.986	0.959	0.917	0.863	0.801
20	0.996	0.988	0.972	0.949	0.917

Moment conditions $E(z_i u_i) = 0$ are invalid.
Simulated probabilities from $1,000,000$ draws of $\chi^2_{k_z-k_x}\chi^2_{k_z-k_x+1}/\chi^2_d$

For the individual 2SLS based tests $S\left(\widehat{\delta}_j\right)$ we also get a rejection frequency less than nominal size for those models where the moment restrictions $E(z_i\varepsilon_{ji}) = 0$ hold, and a dilution of power for those where the moment restrictions do not hold. However, it is well known that the 2SLS based Sargan test is sensitive to weak identification, see e.g. Staiger and Stock (1997), who found that $S\left(\widehat{\beta}_{2sls}\right)$ could severely overreject a true null of $E(z_i u_i) = 0$ in weakly identified models when there was a very strong correlation between the u_i and v_i. This was not the case for $S\left(\widehat{\beta}_L\right)$. Weak identification for β is when Π is near to a rank reduction of 1. For the underidentification test, weak identification for δ means that Π is near a rank reduction of 2. It is therefore informative to compute both the invariant and non-invariant statistics for both over- and underidentification tests. The same holds for robust test statistics, where the CU-GMM based tests have been shown by Newey and Windmeijer (2009) to be well behaved under weak identification, see also Hausman et al. (2012) and Chao et al. (2014). These authors further show that the LIML estimator can have a severe bias in weakly identified models with heteroskedasticity. An invariant jackknife version, HLIM, is shown to be better behaved in that case, which could then be an alternative estimator to use for the robust invariant score tests for over- and underidentification.

9 Asset-Pricing Models

Gospodinov, Kan and Robotti (2017) (GKR) considered the behaviour of the CU-GMM J-test in a reduced-rank asset-pricing model. Using their notation, the candidate stochastic discount factor (SDF) at time t is $x_t'\lambda$, for $t = 1, ..., T$, where $x_t = \begin{pmatrix} 1 & f_t' \end{pmatrix}'$, with f_t a k_f vector of systematic risk factors, and $\lambda = \begin{pmatrix} \lambda_0 & \lambda_1' \end{pmatrix}'$ a k_x-vector of SDF parameters, with $k_x = k_f + 1$. r_t is the k_r-vector of gross returns on $k_r > k_x$ test assets. Let ι_{k_r} be a k_r-vector of ones, then the moment conditions to be tested are given by

$$E\left(r_t x_t'\lambda - \iota_{k_r}\right) = 0. \tag{36}$$

Let $e_t(\lambda) = r_t x_t'\lambda - 1$ then the CU-GMM J-test is given by

$$J\left(\widehat{\lambda}_{cu}\right) = T\min_{\lambda} \overline{e}(\lambda)' \widehat{V}_e(\lambda)\, \overline{e}(\lambda),$$

where $\overline{e}(\lambda) = \frac{1}{T}\sum_{t=1}^{T} e_t(\lambda)$ and $\widehat{V}_e(\lambda)$ is a consistent estimator of the long-run variance matrix of the sample pricing errors $V_e(\lambda)$.

Let $H = \begin{bmatrix} \iota_{k_r} & E\left(r_t x_t'\right) \end{bmatrix}$, then (36) can be written as $H\psi = 0$, with $\psi = \begin{pmatrix} -1 & \lambda' \end{pmatrix}'$, and GKR showed that therefore the $J\left(\widehat{\lambda}_{cu}\right)$ test for testing moment conditions (36) is equivalent to testing $H_0 : \mathrm{r}(H) = k_x$ against $H_1 : \mathrm{r}(H) = k_x + 1$. From this it follows that if $\mathrm{r}\left(E\left(r_t x_t'\right)\right) < k_x$, the CU-GMM J-test has no power to detect violations of the moment conditions (36). Let P_1 be the $k_r \times (k_r - 1)$ orthonormal matrix whose columns are orthogonal to ι_{k_r}, such that $P_1'P_1 = I_{k_r-1}$ and $P_1P_1' = I_{k_r} - \iota_{k_r}\iota_{k_r}'/k_r$. Let $r_{1t} = P_1'r_t$, then GKR show that the $J\left(\widehat{\lambda}_{cu}\right)$-test is the same as the robust Cragg-Donald test for testing $H_0 : \mathrm{r}\left(E\left(r_{1t}x_t'\right)\right) = k_x - 1$ against $H_1 : \mathrm{r}\left(E\left(r_{1t}x_t'\right)\right) = k_x$.

The latter formulation fits the testing procedure described in Section 7. The CD test for $H_0 : \mathrm{r}\left(E\left(r_{1t}x_t'\right)\right) = k_x - 1$ is the test for $H_0 : \mathrm{r}\left(\Lambda_1'\right) = k_x - 1$ in the regression model

$$r_{1t} = \Lambda_1' x_t + v_t,$$

based on the OLS estimator of the $k_x \times (k_r - 1)$ matrix Λ_1. The CD and KP tests are therefore the same as for the tests of $H_0 : \mathrm{r}(C_1) = k_x - 1$ in the specification

$$x_t = C_1' r_{1t} + u_t.$$

With $X = \left[\begin{array}{cc} \iota_T & F \end{array} \right]$, we can therefore obtain robust invariant tests for overidentification by for example estimating the specification

$$\iota_T = F\delta + \varepsilon \tag{37}$$

by LIML, using the $T \times (k_r - 1)$ matrix R_1 as instruments, and computing the Kleibergen-Paap robust score test $S_r\left(\widehat{\delta}_L\right)$, the $J\left(\widehat{\delta}_{2L}\right)$ or $J\left(\widehat{\delta}_{2L,r}\right)$ tests. These invariant robust tests are alternatives to the CU-GMM $J\left(\widehat{\delta}_{cu}\right)$-test that do not have the problems associated with the CU-GMM estimator, which is often more difficult to compute and may give rise to multiple local minima, see the discussion in Peñaranda and Sentana (2015), who argue strongly for the use of invariant methods in these models. For the same reason, Burnside (2016) used the KP test instead of the robust CD test in his simulations.

Note that the assumptions and standardisation used in Theorem 2 in GKR to obtain the limiting distribution of the overidentification test is equivalent to LIML applied to (37), hence the equivalence of their result and the one reported for the Sargan test in the previous section. A conservative upper bound for the CU-GMM J-test is derived in Theorem 1 of GKR. If the model is underidentified, then $\lim_{T\to\infty} \Pr\left(J\left(\widehat{\lambda}_{cu}\right) \leq a\right) = \lim_{T\to\infty} \Pr\left(J\left(\widehat{\delta}_{cu}\right) \leq a\right) \geq \Pr\left(c_{k_r-1} \leq a\right)$, where $c_{k_r-1} \sim \chi^2_{k_r-1}$.

GKR confirmed in a Monte Carlo study the poor performance of the overidentification test in underidentified models. Here, the model is underidentified if $\mathrm{r}\left(E\left(r_t x_t'\right)\right) < k_x$ and GKR incorporated underidentification by including spurious factors that are uncorrelated with the test assets. They did, however, not perform tests of underidentification on the rank of $E\left(r_t x_t'\right)$ itself. These tests are the same as the overidentification tests above for model (37), but now with the $T \times k_r$ matrix R as instruments instead of R_1.

Table 3 presents some results of both the over- and underidentification tests for the same model design as the Monte Carlo exercise in GKR, (Table 1, p 1621). We focus on the misspecified model with 3 useful factors for the cases of 0,1 and 2 spurious factors, for the sample sizes $T = 200, 1000$. The DGP in GKR is one of homoskedastic i.i.d. data, and the robustness they consider in the estimation is that against conditional heteroskedasticity. We repeat this design, including the non-robust Sargan test $S\left(\widehat{\delta}_L\right)$ test as well for comparison and focus only on the Sargan versions of the tests. Unlike GKR, we do not take deviations from the means of the residuals when computing the variance-covariance matrix of the moments.

As in GKR, the results in Table 3 confirm the poor power properties of the test for overidentifying restrictions in underidentified models. The limiting distribution results for the invariant overidentification test as derived by GKR also apply to the invariant underidentification test, meaning that the test statistics converge in distribution to a $\chi_{k_r-k_x+1}$ under the null that $\mathrm{r}\left(E\left(r_t x_t'\right)\right)=k_x-1$ and the maintained assumptions. If $\mathrm{r}\left(E\left(r_t x_t'\right)\right)<k_x-1$ then the rejection frequency of the test will be less than the size of the test. This is confirmed in the results below. In this design, the underidentification tests correctly convey that the model is underidentified. Table 3 also present the results for the two-step GMM Hansen J-test, $S_r\left(\widehat{\lambda}_{2sls}\right)$, confirming that this test does retain power to reject the false null in underidentified models.

Table 4 presents the results for the robust individual Hansen J-tests $S_r\left(\widehat{\delta}_j\right)$, for the models $x_j=X_{-j}\delta_j+\varepsilon_j$, estimated by 2SLS. For this design, these tests give a clear indication of which factors are the spurious ones.

Table 3. Rejection frequencies of over- (R_1) and under- (R) identification tests at 5% level

T	# spur factors	"inst"	$S\left(\widehat{\delta}_L\right)$	$S_r\left(\widehat{\delta}_L\right)$	$J\left(\widehat{\delta}_{2L}\right)$	$J\left(\widehat{\delta}_{cu}\right)$	$S_r\left(\widehat{\lambda}_{2sls}\right)$
200	0	R_1	0.4848	0.4309	0.4234	0.3448	0.5604
		R	0.9991	0.9840	0.9826	0.9778	
	1	R_1	0.0117	0.0112	0.0092	0.0019	0.4793
		R	0.0396	0.0275	0.0241	0.0138	
	2	R_1	0.0002	0.0005	0.0004	0.0000	0.3909
		R	0.0013	0.0015	0.0011	0.0002	
1000	0	R_1	1.0000	1.0000	1.0000	1.0000	1.0000
		R	1.0000	1.0000	1.0000	1.0000	
	1	R_1	0.0424	0.0382	0.0377	0.0371	0.9808
		R	0.0487	0.0433	0.0429	0.0428	
	2	R_1	0.0012	0.0008	0.0008	0.0007	0.9309
		R	0.0016	0.0018	0.0017	0.0012	

Table 4. Rejection frequencies of $S_r\left(\widehat{\delta}_j\right)$ at 5% level

T	# spur factors	$S_r\left(\widehat{\delta}_\iota\right)$	$S_r\left(\widehat{\delta}_{f_1}\right)$	$S_r\left(\widehat{\delta}_{f_2}\right)$	$S_r\left(\widehat{\delta}_{f_3}\right)$	$S_r\left(\widehat{\delta}_{f_4}\right)$	$S_r\left(\widehat{\delta}_{f_5}\right)$
200	0	1.0000	0.9870	0.9998	0.9999		
	1	0.9955	0.9257	0.9810	0.9862	0.0303	
	2	0.9802	0.8319	0.9446	0.9442	0.0263	0.0242
1000	0	1.0000	1.0000	1.0000	1.0000		
	1	0.9951	0.9947	0.9963	0.9961	0.0432	
	2	0.9879	0.9797	0.9835	0.9827	0.0353	0.0346

We can apply the results developed in the previous sections also to the setting of Manresa, Peñaranda and Sentana (2017) (MPS) who considered asset-pricing model moments of the form

$$E\left(r_t x_t'\theta\right) = 0.$$

Maintaining that $E\left(r_t\right) \neq 0$, let $\mathrm{r}\left(E\left(r_t x_t'\right)\right) = k_f + 1 - d$. When $d \geq 2$, there will be a multidimensional subspace of admissible SDFs even after fixing their scale, and MPS, following Arellano, Sentana and Hansen (2012), proceed by estimating a basis of that subspace by replicating d times the moment conditions:

$$E\begin{bmatrix} r_t x_t'\theta_1 \\ r_t x_t'\theta_2 \\ \vdots \\ r_t x_t'\theta_d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

imposing enough normalisation on the parameters to ensure point identification.

For example, for $d = 2$, MPS consider in their application with a model with three factors, the following extended moments

$$E\begin{bmatrix} r_t\left(1 - \mathbf{f}_{12,t}'\delta_1\right) \\ r_t\left(1 - \mathbf{f}_{13,t}'\delta_2\right) \end{bmatrix} = 0, \tag{38}$$

where

$$\mathbf{f}_{12,t}' = \begin{pmatrix} f_{1t} & f_{2t} \end{pmatrix};\ \mathbf{f}_{13,t}' = \begin{pmatrix} f_{1t} & f_{3t} \end{pmatrix}.$$

They proceed to estimate the parameters δ_1 and δ_2 by CU-GMM to obtain $J\left(\widehat{\delta}_{cu}\right)$ as an underidentification test. As the CU-GMM estimator is invariant to normalisation, the same test result is obtained from specifying the moment conditions as

$$E\begin{bmatrix} r_t\left(1 - \mathbf{f}_{23,t}'\delta_1\right) \\ r_t\left(f_{1t} - \mathbf{f}_{23t}'\delta_2\right) \end{bmatrix} = 0,$$

where $\mathbf{f}'_{23,t} = \begin{pmatrix} f_{2t} & f_{3t} \end{pmatrix}$. Therefore this test is the same as the robust CD test for testing $H_0 : \mathrm{r}(C) = k_x - 2$ in

$$x_t = C' r_t + u_t,$$

following the exposition in Section 7 and the general robust CD, CU-GMM rank test as described in Section 6.

10 Conclusions

This paper has developed the links between overidentification tests, underidentification tests, score tests and the Cragg-Donald and Kleibergen-Paap rank tests. This general framework made it possible to establish valid robust underidentification tests for models where these have not been proposed before, like dynamic panel data models estimated by GMM. It is well known that these models may suffer from weak instrument problems, and the example we examined for illustration did indicate that the model was underidentified. An issue with robust underidentification tests is that there is no longer a link with testing for weak instruments as in Stock and Yogo (2005). Therefore, a rejection of the null does not necessarily imply strong instruments. However, if the null of underidentification is not rejected, this clearly suggests a problem with the identification of the model. Also, if an invariant rank test is used for a test for overidentifying restrictions, this test will not have power to reject a false null if the model is underidentified. Given the different behaviours of these test statistics in under- and weakly identified models, it is recommended to calculate invariant and non-invariant over- and underidentification tests for each application.

As an avenue for future research, it is important to establish the behaviour of the tests in weakly identified models, including those with many instruments. The CU-GMM tests are relatively insensitive to weak identification, see Newey and Windmeijer (2009). For cross-sectional models with heteroskedasticity, the proposal of Chao et al. (2014) using Jackknife LIML (HLIM) or Fuller (HFUL) together with their proposed $\widehat{T}$ statistic for overidentification appears a promising avenue, also for testing underidentification as this is simply applying the $\widehat{T}$ test to the linear auxiliary model.

Another issue for future research is the behaviour of the iterated CU-GMM estimator as proposed in Section 3.4. For example, if starting from the 2SLS estimator, how does the transition to the CUE estimator over the iterations develop and after how many iterations does this estimator establishes CU-GMM like properties? One could of course also start the iteration process from HLIM or HFUL and for example investigate the behaviour of the estimator and J-test after one iteration. Another interesting issue is how the iterated CU-GMM estimator behaves in weakly identified models.

References

Alonso-Borrego, C. and M. Arellano, 1999. Symmetrically normalized instrumental-variable estimation using panel data. *Journal of Business & Economic Statistics* 17, 36-49.

Al-Sadoon, M., 2017, A unifying theory of tests of rank. *Journal of Econometrics* 199, 49-62.

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351.

Andrews, I., 2017, Valid two-step identification-robust confidence sets for GMM. *Review of Economics and Statistics*, forthcoming, http://dx.doi.org/10.1162/REST_a_00682.

Arellano, M. and S. Bond, 1998. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, 277-298.

Arellano, M., L.P. Hansen and E. Sentana, 2012. Underidentification? *Journal of Econometrics* 170, 256-280.

Basmann, R.L., 1960. On finite sample distributions of generalized classical linear identifiability test statistics. *Journal of the American Statistical Association* 55, 650-59.

Baum, C.F., M.E. Schaffer and S. Stillman, 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal* 7, 465-506.

Baum, C.F., M.E. Schaffer and S. Stillman, 2010. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Bazzi, S. and M.A. Clemens, 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*

5, 152-186.

Blundell, R. and S. Bond, 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

Blundell, R. and S. Bond, 2000. GMM estimation with persistent panel data: An application to production functions. *Econometric Reviews* 19, 321-340.

Bowden, R.J. and D.A. Turkington, 1984. *Instrumental Variables.* Econometric Society Monographs in Quantitative Economics, Cambridge: Cambridge University Press.

Bun, M.J.G. and M. de Haan, 2010. Weak instruments and the first stage F-statistic in IV models with a nonscalar error covariance structure. Discussion Paper 2010/02, UvA Econometrics, University of Amsterdam.

Bun, M.J.G. and F. Windmeijer, 2010. The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal* 13, 95-126.

Burnside, C., 2016. Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics* 14, 295-330.

Chao, J.C., J.A. Hausman, W.K. Newey, N.R. Swanson and T. Woutersen, 2014. Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics* 178, 15-21.

Chernozhukov, V. and C. Hansen, 2008, The reduced form: A simple approach to inference with weak instruments. *Economics Letters 100*, 68-71.

Cragg, J.G. and S.G. Donald, 1993. Testing identifiability and specification in instrumental variables models. *Econometric Theory* 9, 222-240.

Cragg, J.G. and S.G. Donald, 1997. Inferring the rank of a matrix. *Journal of Econometrics* 76, 223-250.

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

Godfrey, L.G., and M.R. Wickens, 1982. A simple derivation of the limited information maximum likelihood estimator. *Economics Letters* 10, 277-283.

Gospodinov, N., R. Kan and C. Robotti, 2017. Spurious inference in reduced-rank asset-pricing models. *Econometrica* 85, 1613-1628.

Hausman, J.A., W.K. Newey, T. Woutersen, J.C. Chao and N.R. Swanson, 2012. Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3, 211-255.

Hayakawa, K. and M. Qi, 2017. Further results on the weak instruments problem of the system GMM estimator in dynamic panel data models. working paper, http://dx.doi.org/10.2139/ssrn.2965281

Kim, U., 2017, Weak instruments and the robust first-stage F-statistic in an IV regression with heteroskedastic errors. *Mimeo*, Dept. of Economics, University of Bristol.

Kitamura, Y., 2005, Specification tests with instrumental variables and rank deficiency, in D. Corbae, S.N. Durlauf and B.E. Hansen (Eds.), *Econometric Theory and Practice*, 59-81. New York: Cambridge University Press.

Kleibergen, F. and R. Paap, 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133, 97-126.

Manresa, E., F. Peñaranda and E. Sentana, 2017. Empirical evaluation of overspecified asset pricing models. CEPR Discussion Paper Series No. DP12085, CEPR, London.

Newey, W.K., and F. Windmeijer, 2009. Generalized Method of Moments with many weak moment conditions. *Econometrica* 77, 687-719.

Peñaranda, F. and E. Sentana, 2015. A unifying approach to the empirical evaluation of asset pricing models, *The Review of Economics and Statistics* 97, 412-435.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal* 9, 86-136.

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

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

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

Stock, J.H. and M. Yogo, 2005. Testing for weak instruments in linear IV regression. In D.W.K. Andrews and J.H. Stock (Eds.), *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, 80-108. New York: Cambridge University Press.

Wooldridge, J.M., 1995. Score diagnostics for linear models estimated by two stage least squares. In G.S. Maddala, P.C.B. Phillips and T.N. Srinivasan (Eds.), *Advances in Econometrics and Quantitative Economics, Essays in honor of Professor C.R. Rao*, 66-87. Oxford: Blackwell Publishers.

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data. Second Edition.*The MIT Press, Cambridge and London.

Appendix

Proof of Proposition 1. Consider the score test for $H_0 : \gamma = 0$ in model (3),

$$\begin{aligned} y &= X\beta + Z_2\gamma + u \\ &= D\theta + u, \end{aligned}$$

with $D = \begin{bmatrix} X & Z_2 \end{bmatrix}$ and $\theta = \begin{pmatrix} \beta' & \gamma' \end{pmatrix}'$. The full instrument matrix is $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$. The null hypothesis can therefore be written as $H_0 : R\theta = 0$, with $R = \begin{bmatrix} O_{k_x} & I_{k_z - k_x} \end{bmatrix}$, where O_{k_x} is a $(k_z - k_x) \times k_x$ matrix of zeros. As the unrestricted model is just identified, the score for all IV estimators in the unrestricted model is given by

$$s\left(\widehat{\theta}\right) = Z'\left(y - D\widehat{\theta}\right) = 0,$$

with

$$\begin{aligned} \widehat{\theta} &= (Z'D)^{-1} Z'y \\ V\widehat{ar}\left(\widehat{\theta}\right) &= \left(D'Z\widehat{\Omega}_{z\widehat{u}}^{-1}Z'D\right)^{-1} = (Z'D)^{-1}\widehat{\Omega}_{z,\widehat{u}}(D'Z)^{-1}, \end{aligned}$$

where here $\widehat{u} = y - D\widehat{\theta}$.

Let $\widehat{\beta}_1$ be any one-step GMM estimator of β in the restricted model, and let $\widehat{u}_1 = y - X\widehat{\beta}_1$. Then the robust score test statistic for testing the null $H_0 : \gamma = 0$ is given by

$$S_r\left(\widehat{\beta}_1\right) = \widehat{u}_1'Z(D'Z)^{-1}R'\left(R(Z'D)^{-1}\widehat{\Omega}_{z,\widehat{u}_1}(D'Z)^{-1}R'\right)^{-1}R(Z'D)^{-1}Z'\widehat{u}_1.$$

$\widehat{\Omega}_{z,\widehat{u}_1}$ can be written as $Z'H_{\widehat{u}_1}Z$, for example for a cross-sectional heteroskedastic robust estimator, $H_{\widehat{u}_1} = \text{diag}\left(\widehat{u}_{1i}^2\right)$. As

$$(Z'D)^{-1}Z' = \left(\widehat{D}'\widehat{D}\right)^{-1}\widehat{D}'$$

where

$$\widehat{D} = P_Z D = \begin{bmatrix} P_Z X & Z_2 \end{bmatrix} = \begin{bmatrix} \widehat{X} & Z_2 \end{bmatrix},$$

it follows that

$$S_r\left(\widehat{\beta}_1\right) = \widehat{u}_1'\widehat{D}\left(\widehat{D}'\widehat{D}\right)^{-1}R'\left(R\left(\widehat{D}'\widehat{D}\right)^{-1}\widehat{D}'H_{\widehat{u}_1}\widehat{D}\left(\widehat{D}'\widehat{D}\right)^{-1}R'\right)^{-1}R\left(\widehat{D}'\widehat{D}\right)^{-1}\widehat{D}'\widehat{u}_1.$$

As

$$R\left(\widehat{D}'\widehat{D}\right)^{-1}\widehat{D}' = \left(Z_2'M_{\widehat{X}}Z_2\right)^{-1}Z_2'M_{\widehat{X}}$$

it follows that

$$S_r\left(\widehat{\beta}_1\right) = \widehat{u}_1'M_{\widehat{X}}Z_2\left(Z_2'M_{\widehat{X}}H_{\widehat{u}_1}M_{\widehat{X}}Z_2\right)^{-1}Z_2'M_{\widehat{X}}\widehat{u}_1.$$

But

$$\begin{aligned} Z_2'M_{\widehat{X}}\widehat{u}_1 &= Z_2'y - Z_2'X\widehat{\beta}_1 - Z_2'P_{\widehat{X}}y + Z_2'P_{\widehat{X}}X\widehat{\beta}_1 \\ &= Z_2'M_{\widehat{X}}y, \end{aligned}$$

and so we obtain

$$\begin{aligned} S_r\left(\widehat{\beta}_1\right) &= y'M_{\widehat{X}}Z_2\left(Z_2'M_{\widehat{X}}H_{\widehat{u}_1}M_{\widehat{X}}Z_2\right)^{-1}Z_2'M_{\widehat{X}}y \\ &= y'M_{\widehat{X}}Z\widehat{\Omega}^{-1}_{\widetilde{z}_2,\widehat{u}_1}Z_2'M_{\widehat{X}}y\,. \end{aligned}$$

Next, let $\widehat{\beta}_2$ be the two-step GMM estimator, and consider the following version of the robust score test

$$\begin{aligned} S_r\left(\widehat{\beta}_2,\widehat{\beta}_1\right) &= \widehat{u}_2'Z\left(D'Z\right)^{-1}R'\left(R(Z'D)^{-1}\widehat{\Omega}_{z,\widehat{u}_1}\left(D'Z\right)^{-1}R'\right)^{-1}R\left(Z'D\right)^{-1}Z'\widehat{u}_2 \\ &= \widehat{u}_2'M_{\widehat{X}}Z_2\left(Z_2'M_{\widehat{X}}H_{\widehat{u}_1}M_{\widehat{X}}Z_2\right)^{-1}Z_2'M_{\widehat{X}}\widehat{u}_2. \end{aligned}$$

As

$$Z_2'M_{\widehat{X}}\widehat{u}_2 = Z_2'M_{\widehat{X}}y,$$

it follows that

$$S_r\left(\widehat{\beta}_2,\widehat{\beta}_1\right) = S_r\left(\widehat{\beta}_1\right).$$

The score of the two-step estimator in the restricted model is $X'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'\widehat{u}_2 = 0$ and hence $L'D'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'\widehat{u}_2 = 0$, where $L' = \left[\begin{array}{cc} I_{k_x} & O_{k_z-k_x}\end{array}\right]$. As

$$\left(Z'D\right)^{-1}Z'\widehat{u}_2 = \left(D'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'D\right)^{-1}D'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'\widehat{u}_2,$$

and letting $B = D'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'D$, we get

$$S_r\left(\widehat{\beta}_2,\widehat{\beta}_1\right) = \widehat{u}_2'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'DB^{-1}R'\left(RB^{-1}R'\right)^{-1}RB^{-1}D'Z\widehat{\Omega}^{-1}_{z,\widehat{u}_1}Z'\widehat{u}_2.$$

Because $RL = 0$, it follows that, see e.g. Wooldridge (2010, p. 424),

$$\begin{aligned} & B^{-1}R'\left(RB^{-1}R'\right)^{-1}RB^{-1} \\ = \; & B^{-1} - L\left(L'BL\right)^{-1}L', \end{aligned}$$

and so

$$\begin{aligned} S_r\left(\widehat{\beta}_2, \widehat{\beta}_1\right) &= \widehat{u}_2' Z \widehat{\Omega}_{z,\widehat{u}_1}^{-1} Z'D\left(D'Z\widehat{\Omega}_{z,\widehat{u}_1}^{-1}Z'D\right)^{-1} D'Z\widehat{\Omega}_{z,\widehat{u}_1}^{-1}Z'\widehat{u}_2 \\ &= \widehat{u}_2' Z \widehat{\Omega}_{z,\widehat{u}_1}^{-1} Z'\widehat{u}_2 \\ &= J\left(\widehat{\beta}_2, \widehat{\beta}_1\right) \end{aligned}$$

where $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right)$ is the GMM Hansen J-test for overidentifying restrictions. ■

Proof of Proposition 2. It is illustrative to first set $G = I_{k_z}$ and $F = I_{k_x+1}$, hence $\Theta = \Pi^*$ and $\widehat{\Theta} = \widehat{\Pi}^*$. Order the columns of Π^* such that $\Pi^* = \left[\begin{array}{cc} \Pi & \pi_y \end{array}\right]$, and likewise for $\widehat{\Pi}^*$. It then follows from the discussion in Kleibergen and Paap (2006, pp. 101-102) for the IV model, that for $q = k_x$ and $\pi_y = \Pi\beta$

$$\begin{aligned} A_q &= \Pi; B_q = \left[\begin{array}{cc} I_{k_x} & \beta \end{array}\right] \\ A_q B_q &= \left[\begin{array}{cc} \Pi & \Pi\beta \end{array}\right] \\ A_{q,\perp} &= \left(\begin{array}{c} -\left(\Pi_1'\right)^{-1}\Pi_2' \\ I_{k_z-k_x} \end{array}\right)\left(I_d + \Pi_2\Pi_1^{-1}\left(\Pi_1'\right)^{-1}\Pi_2'\right)^{-1/2} \\ B_{q,\perp} &= \left(\begin{array}{cc} 1 & -\beta' \end{array}\right)/\sqrt{1+\beta'\beta} = \psi'/\sqrt{\psi'\psi}. \end{aligned}$$

For the test statistic, we can ignore the standardisation terms $\left(I_{k_z-k_x} + \Pi_2\Pi_1^{-1}\left(\Pi_1'\right)^{-1}\Pi_2'\right)^{-1/2}$ and $\left(\psi'\psi\right)^{-1/2}$. It then follows that the test statistic is based on

$$\widehat{\Lambda}_q = \left[\begin{array}{cc} -\widetilde{\Pi}_2\widetilde{\Pi}_1^{-1} & I_{k_z-k_x} \end{array}\right]\widehat{\Pi}^*\widetilde{\psi} \tag{A.1}$$

where the estimators $\widetilde{\Pi}$ and $\widetilde{\Pi}\widetilde{\beta}$ are determined from

$$\widehat{A}_q\widehat{B}_q = \left[\begin{array}{cc} \widetilde{\Pi} & \widetilde{\Pi}\widetilde{\beta} \end{array}\right].$$

We therefore see that $\widehat{\Lambda}_q$ has the same formula as the estimators for γ in (9) and (16), given the estimates for Π and β. For this case where $G = I_{k_z}$ and $F = I_{k_x+1}$, $\widetilde{\Pi}$ and $\widetilde{\beta}$ are given by

$$\left(\widetilde{\Pi}, \widetilde{\beta}\right) = \arg\min_{\beta,\Pi}\left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \Pi\beta \\ \pi \end{array}\right)\right)'\left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \Pi\beta \\ \pi \end{array}\right)\right).$$

Exactly the same formula for $\widehat{\Lambda}_q$ as in (A.1) is obtained for general choices of F and G, only the estimators $\widetilde{\Pi}$ and $\widetilde{\beta}$ vary with F and G. Denote these estimators $\widetilde{\Pi}_{GF}$ and $\widetilde{\beta}_{GF}$. Then the decomposition for $\widehat{\Theta}$ is

$$\widehat{\Theta} = G\widehat{\Pi}^* F' = \widehat{A}_q \widehat{B}_q + \widehat{A}_{q,\perp}\widehat{\Lambda}_q \widehat{B}_{q,\perp}$$

and hence

$$\widehat{\Pi}^* = (G'G)^{-1} G' \left(\widehat{A}_q \widehat{B}_q + \widehat{A}_{q,\perp}\widehat{\Lambda}_q \widehat{B}_{q,\perp}\right) F (F'F)^{-1},$$

from which it follows that

$$(G'G)^{-1} G' \left(\widehat{A}_q \widehat{B}_q\right) F (F'F)^{-1} = \left[\begin{array}{cc} \widetilde{\Pi}_{GF} & \widetilde{\Pi}_{GF}\widetilde{\beta}_{GF} \end{array}\right],$$

with

$$\left(\widetilde{\Pi}_{GF}, \widetilde{\beta}_{GF}\right) = \arg\min_{\beta,\Pi} \left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \Pi\beta \\ \pi \end{array}\right)\right)' (F'F \otimes G'G) \left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \Pi\beta \\ \pi \end{array}\right)\right).$$

■

Proof of Lemma 1. Following Bowden and Turkington (1984, pp. 112-113), consider the following minimisation problem

$$\min_{\beta,\Pi^*} \frac{1}{2} \left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \pi_y \\ \pi \end{array}\right)\right)' (A \otimes B) \left(\left(\begin{array}{c} \widehat{\pi}_y \\ \widehat{\pi} \end{array}\right) - \left(\begin{array}{c} \pi_y \\ \pi \end{array}\right)\right)$$

$$s.t.\ \Pi\beta - \pi_y = \Pi^*\psi = 0$$

where $\Pi^* = \left[\begin{array}{cc} \pi_y & \Pi \end{array}\right]$ and $\psi = \left(\begin{array}{cc} -1 & \beta' \end{array}\right)'$. A and B are $(k_x+1)\times(k_x+1)$ and $k_z \times k_z$ symmetric nonsingular matrices respectively.

The Lagrangean is given by

$$L(\pi^*, \beta, \mu) = \frac{1}{2} (\widehat{\pi}^* - \pi^*)' (A \otimes B) (\widehat{\pi}^* - \pi^*) + \mu'\Pi^*\psi.$$

Let a tilde $\widetilde{}$ denote the constrained estimators, then the first-order conditions are given by

$$\frac{\partial L(\pi^*, \beta, \mu)}{\partial \pi^*} = -(A \otimes B)(\widehat{\pi}^* - \widetilde{\pi}^*) + \left(\widetilde{\psi} \otimes \widetilde{\mu}\right) = 0 \quad (A.2)$$

$$\frac{\partial L(\pi^*, \beta, \mu)}{\partial \beta} = \widetilde{\Pi}'\widetilde{\mu} = 0 \quad (A.3)$$

$$\frac{\partial L(\pi^*, \beta, \mu)}{\partial \mu} = \widetilde{\Pi}^*\widetilde{\psi} = 0. \quad (A.4)$$

From (A.2) it follows that

$$-B\left(\widehat{\Pi}^{*}-\widetilde{\Pi}^{*}\right)A+\widetilde{\mu}\widetilde{\psi}'=0. \tag{A.5}$$

Hence, postmultiplying by $A^{-1}\widetilde{\psi}$

$$-B\left(\widehat{\Pi}^{*}-\widetilde{\Pi}^{*}\right)\widetilde{\psi}+\widetilde{\mu}\widetilde{\psi}'A^{-1}\widetilde{\psi}=0$$

and so

$$\widetilde{\mu}=\frac{B\left(\widehat{\Pi}^{*}-\widetilde{\Pi}^{*}\right)\widetilde{\psi}}{\widetilde{\psi}'A^{-1}\widetilde{\psi}}=\frac{B\widehat{\Pi}^{*}\widetilde{\psi}}{\widetilde{\psi}'A^{-1}\widetilde{\psi}}.$$

From (A.3) it then follows that

$$\widetilde{\Pi}'\widetilde{\mu}=\frac{\widetilde{\Pi}'B\widehat{\Pi}^{*}\widetilde{\psi}}{\widetilde{\psi}'A^{-1}\widetilde{\psi}}=0,$$

and so $\widetilde{\Pi}B\left(\widehat{\pi}_{y}-\widehat{\Pi}\widetilde{\beta}\right)=0$, or

$$\widetilde{\beta}=\left(\widetilde{\Pi}'B\widehat{\Pi}\right)^{-1}\widetilde{\Pi}'B\widehat{\pi}_{y}. \tag{A.6}$$

Therefore, if $B=Z'Z$, then

$$\begin{aligned}\widetilde{\beta} &= \left(\widetilde{\Pi}Z'Z\widehat{\Pi}\right)^{-1}\widetilde{\Pi}Z'Z\widehat{\pi}_{y}\\ &= \left(\widetilde{X}'\widehat{X}\right)^{-1}\widetilde{X}'y=\left(\widetilde{X}'X\right)^{-1}\widetilde{X}'y.\end{aligned}$$

Further, from (A.5)

$$\widetilde{\Pi}^{*}=\widehat{\Pi}^{*}-B^{-1}\widetilde{\mu}\widetilde{\psi}'A^{-1}$$

it follows that, with $C=\left[\begin{array}{cc}0 & I_{k_x}\end{array}\right]'$,

$$\begin{aligned}\widetilde{\Pi} &= \widehat{\Pi}-\frac{\widehat{\Pi}^{*}\widetilde{\psi}\widetilde{\psi}'A^{-1}C}{\widetilde{\psi}'A^{-1}\widetilde{\psi}}\\ &= \widehat{\Pi}-\frac{\left(\widehat{\Pi}\widetilde{\beta}-\widehat{\pi}_{y}\right)\widetilde{\psi}\widetilde{\psi}'A^{-1}C}{\widetilde{\psi}'A^{-1}\widetilde{\psi}}\end{aligned}$$

Therefore

$$\widetilde{\Pi}'B\widetilde{\Pi}=\widetilde{\Pi}'B\widehat{\Pi}-\frac{\widetilde{\Pi}'B\left(\widehat{\Pi}\widetilde{\beta}-\widehat{\pi}_{y}\right)\widetilde{\psi}'A^{-1}C}{\widetilde{\psi}'A^{-1}\widetilde{\psi}},$$

and so it follows from (A.6) that $\widetilde{\Pi}'B\left(\widehat{\Pi}\widetilde{\beta}-\widehat{\pi}_y\right)=0$, and hence $\widetilde{\Pi}'B\widetilde{\Pi}=\widetilde{\Pi}'B\widehat{\Pi}$. Therefore, when $B=Z'Z$ we get that $\widetilde{\Pi}'Z'Z\widetilde{\Pi}=\widetilde{\Pi}'Z'Z\widehat{\Pi}$ and hence

$$\widetilde{\beta}=\left(\widetilde{X}'\widetilde{X}\right)^{-1}\widetilde{X}'y.$$

■

Proof of Lemma 2. Next consider the following minimisation problem

$$\min_{\beta,\Pi^*}\frac{1}{2}\left(\left(\begin{array}{c}\widehat{\pi}_y\\ \widehat{\pi}\end{array}\right)-\left(\begin{array}{c}\pi_y\\ \pi\end{array}\right)\right)'\left(I_{k_x+1}\otimes B\right)A\left(I_{k_x+1}\otimes B\right)\left(\left(\begin{array}{c}\widehat{\pi}_y\\ \widehat{\pi}\end{array}\right)-\left(\begin{array}{c}\pi_y\\ \pi\end{array}\right)\right)$$

$$s.t.\ \Pi\beta-\pi_y=\Pi^*\psi=0$$

where as above $\Pi^*=\left[\begin{array}{cc}\pi_y & \Pi\end{array}\right]$ and $\psi=\left(\begin{array}{cc}-1 & \beta'\end{array}\right)'$. A and B are $(k_x+1)\,k_z\times(k_x+1)\,k_z$ and $k_z\times k_z$ symmetric nonsingular matrices respectively.

The Lagrangean is given by

$$L\left(\pi^*,\beta,\mu\right)=\frac{1}{2}\left(\widehat{\pi}^*-\pi^*\right)'\left(I_{k_x+1}\otimes B\right)A\left(I_{k_x+1}\otimes B\right)\left(\widehat{\pi}^*-\pi^*\right)+\mu'\Pi^*\psi,$$

and the first-order conditions are given by (A.3), (A.4) and

$$\frac{\partial L\left(\pi^*,\beta,\mu\right)}{\partial\pi^*}=-\left(I_{k_x+1}\otimes B\right)A\left(I_{k_x+1}\otimes B\right)\left(\widehat{\pi}^*-\widetilde{\pi}^*\right)+\left(\widetilde{\psi}\otimes\widetilde{\mu}\right)=0. \tag{A.7}$$

From (A.7) it follows that

$$\begin{aligned}\left(I_{k_x+1}\otimes B\right)\left(\widehat{\pi}^*-\widetilde{\pi}^*\right) &= A^{-1}\left(I_{k_x+1}\otimes B^{-1}\right)\left(\widetilde{\psi}\otimes\widetilde{\mu}\right)\\ &= A^{-1}\left(\widetilde{\psi}\otimes I_{k_z}\right)B^{-1}\widetilde{\mu}\end{aligned}$$

Pre-multiplying both sides by $\left(\widetilde{\psi}'\otimes I_{k_z}\right)$ results in

$$\left(\widetilde{\psi}'\otimes B\right)\left(\widehat{\pi}^*-\widetilde{\pi}^*\right)=\left(\widetilde{\psi}'\otimes I_{k_z}\right)A^{-1}\left(\widetilde{\psi}\otimes I_{k_z}\right)B^{-1}\widetilde{\mu}.$$

As

$$\left(\widetilde{\psi}\otimes B\right)\left(\widehat{\pi}^*-\widetilde{\pi}^*\right)=B\left(\widehat{\Pi}^*-\widetilde{\Pi}^*\right)\widetilde{\psi}=B\widehat{\Pi}^*\widetilde{\psi},$$

it follow that

$$\widetilde{\mu}=B\left(\left(\widetilde{\psi}'\otimes I_{k_z}\right)A^{-1}\left(\widetilde{\psi}\otimes I_{k_z}\right)\right)^{-1}B\widehat{\Pi}^*\widetilde{\psi},$$

and hence the solution for $\widetilde{\beta}$ satisfies

$$\widetilde{\beta}=\left(\widetilde{\Pi}'B\left(\left(\widetilde{\psi}'\otimes I_{k_z}\right)A^{-1}\left(\widetilde{\psi}\otimes I_{k_z}\right)\right)^{-1}B\widehat{\Pi}\right)^{-1}\widetilde{\Pi}'B\left(\left(\widetilde{\psi}'\otimes I_{k_z}\right)A^{-1}\left(\widetilde{\psi}\otimes I_{k_z}\right)\right)^{-1}B\widehat{\pi}_y.$$

Let $B = Z'Z$ and we choose for example $A = \left(\sum_{i=1}^{n}(w_i w_i') \otimes (z_i z_i')\right)^{-1}$ for a heteroskedasticity robust variance estimator of $\widehat{\pi}^*$ under the null that $\pi^* = 0$. We get that $\left(\widetilde{\psi}' \otimes I_{k_z}\right) A^{-1} \left(\widetilde{\psi} \otimes I_{k_z}\right) = \sum_{i=1}^{n} \widetilde{u}_i^2 z_i z_i'$, where $\widetilde{u}_i = w_i \widetilde{\psi} = y_i - x_i' \widetilde{\beta}$, and the solution for the CUE estimator satisfies

$$\widetilde{\beta} = \left(\widetilde{\Pi}' Z' Z \left(\sum_{i=1}^{n} \widetilde{u}_i^2 z_i z_i'\right)^{-1} Z' X\right)^{-1} \widetilde{\Pi}' Z' Z \left(\sum_{i=1}^{n} \widetilde{u}_i^2 z_i z_i'\right)^{-1} Z' y.$$

■