A simple underidentification test for linear IV models, with an application to dynamic panel data models*

Frank Windmeijer†
Department of Economics
University of Bristol, UK
and Cemmap, London, UK

April 2017
Preliminary and Incomplete

Abstract

For linear IV models it is shown that standard underidentification tests like the Cragg-Donald and Kleibergen-Paap tests are Sargan type tests for instrument-error orthogonality in the linear IV model of one endogenous explanatory variable regressed on the others, estimated by LIML. This insight is then used to extend these type of tests to more complex data structures, like dynamic panel data models, leading to very simple to calculate tests for underidentification.

---

*I would like to thank Bo Honoré, Jon Temple and seminar and conference participants at Amsterdam, Bristol, Canterbury, Oxford, Melbourne and Perth for helpful comments.
†E-mail address: f.windmeijer@bristol.ac.uk.

1 Introduction

For linear IV models it is shown that standard underidentification tests like the Cragg-Donald and Kleibergen-Paap tests are Sargan type tests for instrument-error orthogonality in the linear IV model of one endogenous explanatory variable regressed on the others, estimated by LIML. This insight is then used to extend these type of tests to more complex data structures, like dynamic panel data models, leading to very simple to calculate tests for underidentification.

2 Model and Assumptions

Let y_i be a scalar, x_i a k_x-vector and z_i a k_z-vector, with $k_z \geq k_x$. The structural linear model of interest is given by

$$y_i = x_i' \beta + u_i. \tag{1}$$

The explanatory variables x_i are correlated with the unobservables, $E(x_i u_i) \neq 0$, and hence the OLS estimator is biased and inconsistent. We use the instruments z_i for estimation of the parameters β. If we write the model in matrix form

$$y = X\beta + u,$$

where y and u are the n-vectors (y_i) and (u_i), and X is the $n \times k_x$ matrix $[x_i']$, then the IV, or Two-Stage Least Squares (2SLS), estimator is given by

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

where $P_Z = Z(Z'Z)^{-1} Z'$, with Z the $n \times k_z$ matrix $[z_i']$.

We make the following assumptions

Assumption 1 $\{y_i, x_i', z_i'\}_{i=1}^n$ *is an i.i.d. random sample from a population with finite fourth moments.*

Assumption 2 $E(x_i x_i') = Q_{xx}$ *and* $E(z_i z_i') = Q_{zz}$. Q_{xx} *and* Q_{zz} *are nonsingular.*

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

Assumption 4 $E(z_i u_i) = 0$.

Assumption 5 *$E\left(u_i^2 z_i z_i'\right) = \Omega$, a finite and positive definite matrix.*

Under Assumptions 1-5, the commonly used 2SLS estimator is consistent and asymptotically normally distributed, with the limiting distribution given by

$$\sqrt{n}\left(\widehat{\beta}_{2sls} - \beta\right) \xrightarrow{d} N\left(0, V_\beta\right),$$

where $V_\beta = \left(Q_{zx}' Q_{zz}^{-1} Q_{zx}\right)^{-1} Q_{zx}' Q_{zz}^{-1} \Omega Q_{zz}^{-1} Q_{zx} \left(Q_{zx}' Q_{zz}^{-1} Q_{zx}\right)^{-1}$.

Assumption 3 is a necessary condition for the identification of β using the instrumental variables. The first-stage, linear projection model for x_i is given by

$$x_i = \Pi' z_i + v_i, \tag{2}$$

where Π is a $k_z \times k_x$ matrix, and from the linear projection property, $E\left(x_i v_i'\right) = 0$. In matrix form,

$$X = Z\Pi + V,$$

with V the $n \times k_x$ matrix $[v_i']$. Therefore,

$$E\left(z_i x_i'\right) = E\left(z_i z_i' \Pi\right) = Q_{zz}\Pi.$$

For Assumption 3 to be satisfied, $Q_{zz}\Pi$ needs to have full column rank k_x. As Q_{zz} is nonsingular, this means that β is identified iff $\mathrm{r}\left(\Pi\right) = k_x$.

Tests for identification have focused on testing the rank of Π. In the single endogenous variable case, $k_x = 1$, the first-stage model is given by

$$x_i = z_i' \pi + v_i,$$

and the test for under-identification is simply a test for $H_0 : \pi = 0$ against $H_1 : \pi \neq 0$. Standard practice is to report the standard F-test statistic for testing this hypothesis. Staiger and Stock (1997) and Stock and Yogo (2005) extended this to testing for weak instruments using weak instrument asymptotics of the form $\pi = c/\sqrt{n}$, where c is a vector of constants. Their results highlight the fact that the 2SLS estimator may still have substantial bias, and the Wald test for tests on β size distorted, even if $H_0 : \pi = 0$ is rejected at conventional significance levels. Stock and Yogo (2005) provide critical values for the F-test for testing weak instruments hypotheses, see also Skeels and Windmeijer

(2016). The caveat for these weak instrument tests is that they are only valid under conditional homoskedasticity of u_i and v_i, that is

$$E\left(\left(\begin{array}{c} u_i \\ v_i \end{array}\right)\left(\begin{array}{c} u_i \\ v_i \end{array}\right)' |z_i\right) = \Sigma = \left(\begin{array}{cc} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{array}\right),$$

and the derived critical values do not apply to more general settings of conditional heteroskedasticity, see Bun and De Haan (2010) and Andrews (2015).

For multiple endogenous variables, the Stock-Yogo procedure focuses on the Cragg and Donald (1993) minimum eigenvalue test statistic, which is a test for $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$.

This paper considers settings beyond the limited conditional homoskedasticity setup and develops simple testing procedures for underidentification that apply to general linear models. As is standard practice, these tests will focus on testing $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$. A commonly reported robust extension of the Cragg-Donald (CD) test is the Kleibergen-Paap (KP) test, Kleibergen and Paap (2006). We will review both tests below and show that they are (robust) tests for the overidentifying restrictions in the linear IV model of one endogenous variables regressed on the other endogenous variables in the model, estimated by LIML, using the same set of instruments as in the original, structural equation. We call the latter IV regression\model the auxiliary regression\model. As LIML is invariant to normalisation, the CD and KP tests results do not depend on which endogenous explanatory variable has been chosen as the dependent variable in the auxiliary model.

For some models, like linear dynamic panel data models, underidentification test results are not commonly reported. Bazzi and Clemens (2013) advocate use of the KP test, which in this case is a robust test for overidentifyng restrictions in the auxiliary regression based on a pooled LIML estimator. We propose use of simple two-step GMM Hansen tests in the auxiliary models as tests for underidentification, which have the advantage that they can be computed using standard software. These tests are not invariant to normalisation, and are therefore able to provide information on the explanatory power of the instruments per endogenous explanatory variable. This approach can therefore be seen as a robust extension to the conditional F-test procedure of Sanderson and Windmeijer (2016) and Angrist and Pischke (2006).

As the underidentification tests are tests for the overidentifying restrictions in the auxiliary model, we first discuss and derive some new results for testing overidentiying restrictions in the standard structural model (1), estimated by GMM.

3 Tests of Overidentifying Restrictions

The standard tests for overidentifying restrictions, $H_0 : E(z_i u_i) = 0$, when $k_z > k_x$, are the Sargan and Hansen tests. The Sargan test is given by

$$S\left(\widehat{\beta}_{2sls}\right) = \frac{\widehat{u}_{2sls}' Z \left(Z'Z\right)^{-1} Z' \widehat{u}_{2sls}}{\widehat{u}_{2sls}' \widehat{u}_{2sls}/n}$$

and $S\left(\widehat{\beta}_{2sls}\right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under the null and the maintained assumption of conditional homoskedasticity, $E\left(u_i^2 | z_i\right) = \sigma_u^2$ in which case the 2SLS estimator is asymptotically efficient under Assumptions 1 to 5.

Let $\widehat{\Pi} = \left(Z'Z\right)^{-1} Z'X$ and $\widehat{X} = Z\widehat{\Pi}$. Then without loss of generality we can specify the instrument matrix as $Z = \left[\begin{array}{cc} \widehat{X} & Z_2 \end{array}\right]$, where Z_2 is any $k_z - k_x$ subset of instruments. Again wlog, we can set $Z_2 = [z_{k_x+1}, z_{k_x+2}, ..., z_{k_z}]$. As $\widehat{X}'\widehat{u}_{2sls} = 0$, it follows that

$$S\left(\widehat{\beta}_{2sls}\right) = \frac{\widehat{u}_{2sls}' Z_2 \left(Z_2' M_{\widehat{X}} Z_2\right)^{-1} Z_2' \widehat{u}_{2sls}}{\widehat{u}_{2sls}' \widehat{u}_{2sls}/n}. \tag{3}$$

The two-step GMM estimator is asymptotically efficient under general forms of heteroskedasticity. Using $\widehat{\beta}_{2sls}$ as the one-step GMM estimator, let

$$\widehat{\Omega} = \sum_{i=1}^{n} \widehat{u}_{2sls,i}^2 z_i z_i'.$$

Then the two-step GMM estimator is given by

$$\begin{aligned} \widehat{\beta}_{gmm} &= \arg\min_{\beta} \left(y - X\beta\right)' Z\widehat{\Omega}^{-1} Z' \left(y - X\beta\right) \\ &= \left(X'Z\widehat{\Omega}^{-1}Z'X\right)^{-1} X'Z\widehat{\Omega}^{-1}Z'y. \end{aligned}$$

The Hansen J-test for $H_0 : E(z_i u_i) = 0$ is then

$$J\left(\widehat{\beta}_{gmm}\right) = \left(y - X\widehat{\beta}_{gmm}\right)' Z\widehat{\Omega}^{-1} Z' \left(y - X\widehat{\beta}_{gmm}\right) \tag{4}$$

and $J\left(\widehat{\beta}_{gmm}\right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under the null.

Next, let $Z = \left[\begin{array}{cc} Z_1 & Z_2 \end{array} \right]$, with Z_1 and $n \times k_x$ matrix. Partition $\Pi = \left[\begin{array}{cc} \Pi_1' & \Pi_2' \end{array} \right]'$and π_y accordingly. Π_1 is a $k_x \times k_x$ matrix. Then, the first-stage model for y is

$$\begin{aligned} y &= Z\pi_y + v_y = Z_1\pi_{y1} + Z_2\pi_{y2} + v_y \\ &= Z_1\Pi_1\Pi_1^{-1}\pi_{y1} + Z_2\pi_{y2} + v_y \\ &= X\Pi_1^{-1}\pi_{y1} + Z_2\left(\pi_{y2} - \Pi_2\Pi_1^{-1}\pi_{y1}\right) + v_y - V\Pi_1^{-1}\pi_{y1}. \end{aligned}$$

Or

$$y = X\kappa + Z_2\gamma + u, \tag{5}$$

where $\kappa = \Pi_1^{-1}\pi_{y1}$ and $\gamma = \pi_{y2} - \Pi_2\Pi_1^{-1}\pi_{y1}$. Under the null, $H_0 : \pi_y = \Pi\beta$ we have that $\kappa = \beta$ and $\gamma = 0$. The score for the IV estimator in this just identified model, using Z as instruments, is given by

$$Z'\left(y - X\widehat{\beta}_{IV} - Z_2\widehat{\gamma}_{IV}\right) = 0,$$

and the score test for $H_0 : \gamma = 0$ is then based on the vector of contrasts in the restricted model, $Z'\widehat{u}_{2sls}$. Again, using $Z = \left[\begin{array}{cc} \widehat{X} & Z_2 \end{array} \right]$, and so $Z'\widehat{u}_{2sls} = \left[\begin{array}{cc} 0 & \widehat{u}_{2sls}'Z_2 \end{array} \right]'$, we only need to consider the contrasts $Z_2'\widehat{u}_{2sls}$. Let $\widetilde{Z}_2 = M_{\widehat{X}}Z_2$, then

$$\begin{aligned} Z_2'\widehat{u}_{2sls} &= Z_2'\left(I - X\left(\widehat{X}'\widehat{X}\right)\widehat{X}'\right)u \\ &= Z_2'M_{\widehat{X}}u = \widetilde{Z}_2'u, \end{aligned}$$

where the second equality follows from the fact that $Z_2'X = Z_2'P_ZX = Z_2'\widehat{X}$. It then follows under the standard assumptions that

$$\frac{1}{\sqrt{n}}Z_2'\widehat{u}_{2sls} \xrightarrow{d} N\left(0, E\left[u_i^2\widetilde{z}_{2i}\widetilde{z}_{2i}'\right]\right),$$

where $\widetilde{z}_{2i}'$ are the rows of $\widetilde{Z}_2$. A robust score/Sargan test is then given by

$$S_r\left(\widehat{\beta}_{2sls}\right) = \widehat{u}_{2sls}'Z_2\widetilde{\Omega}_2^{-1}Z_2'\widehat{u}_{2sls} \xrightarrow{d} \chi^2_{k_z - k_x}, \tag{6}$$

where

$$\widetilde{\Omega}_2 = \sum_{i=1}^{n} \widehat{u}_{2sls,i}^2\widetilde{z}_{2i}\widetilde{z}_{2i}'.$$

Proposition 1 below shows that the robust Sargan test is the same as the Hansen J-test, $S_r\left(\widehat{\beta}_{2sls}\right) = J\left(\widehat{\beta}_{gmm}\right)$.

The result of the Proposition 1 applies to the general two-step GMM estimation procedure. The one-step GMM estimator 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. For brevity, we have suppressed the dependence of the one-step GMM estimator on the choice of W_n in the notation. Clearly, $\widehat{\beta}_{2sls}$ is the one-step GMM estimator with $W_n = Z'Z$. The one-step residual is given by $\widehat{u}_1 = y - X\widehat{\beta}_1$, and the efficient two-step GMM estimator is

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

where

$$\widehat{\Omega}_{\widehat{\beta}_1} = \sum_{i=1}^{n} \widehat{u}_{1i}^2 z_i z_i'.$$

The Hansen J-test is then

$$J\left(\widehat{\beta}_2\right) = \widehat{u}_2' Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1} Z'\widehat{u}_2, \tag{7}$$

where $\widehat{u}_2 = y - X\widehat{\beta}_2$.

Denote

$$\widehat{X}_w = ZW_n^{-1}Z'X. \tag{8}$$

Then

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

and so $\widehat{X}_w'\widehat{u}_1 = 0$. The score test for $H_0 : \gamma = 0$ in (5) with estimator $\widehat{\beta}_1$ is then based on the vector of contrasts

$$\begin{aligned} Z_2'\widehat{u}_1 &= Z_2'\left(I_{k_z} - X\left(\widehat{X}_w'\widehat{X}\right)^{-1}\widehat{X}_w'\right)u \\ &= Z_2'\left(I_{k_z} - \widehat{X}\left(\widehat{X}_w'\widehat{X}\right)^{-1}\widehat{X}_w'\right)u = \widetilde{Z}_{2w}'u, \end{aligned}$$

where $\widetilde{Z}_{2w} = \left(I_n - \widehat{X}_w\left(\widehat{X}_w'\widehat{X}\right)^{-1}\widehat{X}'\right)Z_2$. Therefore, under homoskedasticity,

$$S\left(\widehat{\beta}_1\right) = \frac{\widehat{u}_1'Z_2\left(\widetilde{Z}_{2w}'\widetilde{Z}_{2w}\right)^{-1}Z_2'\widehat{u}_1}{\widehat{u}_1'\widehat{u}_1/n} \xrightarrow{d} \chi^2_{k_z-k_x}.$$

A robust version of the test is obtained as follows. Let $\widetilde{z}'_{2w,i}$ be the rows of $\widetilde{Z}_{2w}$ and

$$\widetilde{\Omega}_{2w} = \sum_{i=1}^{n} \widehat{u}_{1i}^{2} \widetilde{z}_{2w,i} \widetilde{z}'_{2w,i}.$$

Then

$$S_r\left(\widehat{\beta}_1\right) = \widehat{u}'_1 Z_2 \widetilde{\Omega}_{2w}^{-1} Z'_2 \widehat{u}_1 \xrightarrow{d} \chi^2_{k_z - k_x}. \tag{9}$$

The following proposition states the equivalence of $S_r\left(\widehat{\beta}_1\right)$ and $J\left(\widehat{\beta}_2\right)$.

Proposition 1 *Let* $S_r\left(\widehat{\beta}_1\right)$ *and* $J\left(\widehat{\beta}_2\right)$ *be as defined in (9) and (7) respectively. Then* $S_r\left(\widehat{\beta}_1\right) = J\left(\widehat{\beta}_2\right)$.

Proof. See Appendix. ■

Next, consider the vector of contrasts

$$\begin{aligned}
\widetilde{Z}'_2 \widehat{u}_1 &= Z'_2 M_{\widehat{X}} \widehat{u}_1 \\
&= Z'_2 M_{\widehat{X}} \left(I_{k_z} - X \left(\widehat{X}'_w \widehat{X} \right)^{-1} \widehat{X}'_w \right) u \\
&= Z'_2 M_{\widehat{X}} \left(I_{k_z} - \widehat{X} \left(\widehat{X}'_w \widehat{X} \right)^{-1} \widehat{X}'_w \right) u \\
&= Z'_2 M_{\widehat{X}} u.
\end{aligned}$$

It therefore follows that

$$S_{r,\widehat{x}}\left(\widehat{\beta}_1\right) = \widehat{u}'_1 \widetilde{Z}_2 \widetilde{\Omega}_{21}^{-1} \widetilde{Z}'_2 \widehat{u}_1 \xrightarrow{d} \chi^2_{k_z - k_x}, \tag{10}$$

where

$$\widetilde{\Omega}_{21} = \sum_{i=1}^{n} \widehat{u}_{1i}^{2} \widetilde{z}_{2i} \widetilde{z}'_{2i}.$$

The next proposition shows the equivalence of $S_{r,\widehat{x}}\left(\widehat{\beta}_1\right)$ and $S_r\left(\widehat{\beta}_1\right)$ and hence $J\left(\widehat{\beta}_2\right)$.

Proposition 2 *Let* $S_{r,\widehat{x}}\left(\widehat{\beta}_1\right)$ *and* $S_r\left(\widehat{\beta}_1\right)$ *be as defined in (10) and (9) respectively. Then* $S_{r,\widehat{x}}\left(\widehat{\beta}_1\right) = S_r\left(\widehat{\beta}_1\right)$.

Proof. See Appendix ■

From Proposition 2 it follows that the asymptotic local power of $J\left(\widehat{\beta}_2\right)$, setting $\gamma = c/\sqrt{n}$ in model (5), is not affected by the choice of one-step GMM estimator $\widehat{\beta}_1$, as stated in the next corollary.

Corollary 1 *Consider the local to zero alternative, $\gamma = c/\sqrt{n}$ in model (5), with c a vector of constants. Under this alternative, and using standard asymptotic results, the limiting distribution of $S_{r,\widehat{x}}\left(\widehat{\beta}_1\right)$, and hence $J\left(\widehat{\beta}_2\right)$, is given by*

$$S_{r,\widehat{x}}\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z-k_x}(\lambda),$$

a non-central $\chi^2_{k_z-k_x}$ distributed random variable with non-centrality parameter

$$\lambda = \text{plim}\left(\left(c'M_{\widehat{X}}Z_2/n\right)\left(\widetilde{\Omega}_{21}/n\right)^{-1}\left(Z_2'M_{\widehat{X}}c/n\right)\right),$$

where

$$\text{plim}\left(\frac{1}{n}\widetilde{\Omega}_{21}\right) = E\left[u_i^2\widetilde{z}_{2i}\widetilde{z}_{2i}'\right]. \tag{11}$$

In order to verify (11), let $A_w = \left(I_n - \widehat{X}_w\left(\widehat{X}_w'\widehat{X}\right)^{-1}\widehat{X}'\right)$. Then

$$\widehat{u}_1 = y - X\widehat{\beta}_1 = \left(I - A_w'\right)\left(u + Z_2c/\sqrt{n}\right) = \widetilde{u}_w + \widetilde{Z}_{2w}c/\sqrt{n}.$$

Then

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_{1i}^2\widetilde{z}_{2i}\widetilde{z}_{2i}' = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{u}_{wi}^2 + 2\widetilde{u}_{wi}\widetilde{z}_{2w,i}'c/\sqrt{n} + \widetilde{z}_{2w,i}'c^2/n\right)\widetilde{z}_{2i}\widetilde{z}_{2i}'.$$

The latter two terms converge to 0, as the fourth-order moments are assumed to exist. As

$$\widetilde{u}_{wi} = u_i - \widehat{x}_{wi}'\left(\widehat{X}_w'\widehat{X}\right)^{-1}\sum_{j=1}^{n}\widehat{x}_j u_j,$$

and the data are i.i.d, it follows that

$$\begin{aligned}
\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_{1i}^2\widetilde{z}_{2i}\widetilde{z}_{2i}'\right) &= \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{u}_{wi}^2\widetilde{z}_{2i}\widetilde{z}_{2i}'\right) \\
&= \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}u_i^2\left(1 - \widehat{x}_{wi}'\left(\widehat{X}_w'\widehat{X}/n\right)^{-1}\widehat{x}_i/n\right)\widetilde{z}_{2i}\widetilde{z}_{2i}'\right) \\
&= E\left[u_i^2\widetilde{z}_{2i}\widetilde{z}_{2i}'\right].
\end{aligned}$$

The result of Corollary 1 means that the relative (in)efficiency of $\widehat{\beta}_1$ does not affect the asymptotic power of the Hansen test. An interesting extension is to consider the statistic

$$S_{r,\widehat{x}}\left(\widehat{\beta}_1, \widehat{\beta}_2\right) = \widehat{u}_1'\widetilde{Z}_2\widetilde{\Omega}_{22}^{-1}\widetilde{Z}_2'\widehat{u}_1,$$

where

$$\widetilde{\Omega}_{22}=\sum_{i=1}^{n}\widehat{u}_{2i}^{2}\widetilde{z}_{2i}\widetilde{z}_{2i}'.$$

From the proofs of Propositions 1 and 2, it follows that $S_{r,\widehat{x}}\left(\widehat{\beta}_1,\widehat{\beta}_2\right)$ is equal to the J-test for the three-step GMM estimator. This J-test is given by

$$J\left(\widehat{\beta}_3\right)=\left(y-X\widehat{\beta}_3\right)'Z\left(\widehat{\Omega}_{\widehat{\beta}_2}\right)^{-1}Z'\left(y-X\widehat{\beta}_3\right),$$

where

$$\widehat{\beta}_3=\arg\min_{\beta}\left(y-X\beta\right)'Z\left(\widehat{\Omega}_{\widehat{\beta}_2}\right)^{-1}Z'\left(y-X\beta\right)$$

and

$$\widehat{\Omega}_{\widehat{\beta}_2}=\sum_{i=1}^{n}\widehat{u}_{2i}^{2}z_iz_i'.$$

From the result in Corollary 1 it follows that $J\left(\widehat{\beta}_3\right)$ has the same limiting distribution as $J\left(\widehat{\beta}_2\right)$ under local alternatives, and hence the same asymptotic power.

A way to obtain $S_{r,\widehat{x}}\left(\widehat{\beta}_1\right)$, and hence $J\left(\widehat{\beta}_2\right)$, is as follows. Estimate by OLS the parameter vector η_z^1 in the specification

$$\widehat{u}_1=\widehat{X}\eta_x^1+Z_2\eta_z^1+\xi_1 \tag{12}$$

where the superscripts and subscript denotes the dependence on the choice of $\widehat{\beta}_1$. Denote the estimator by $\widehat{\eta}_z^1$, given by

$$\begin{aligned}\widehat{\eta}_z^1 &= \left(Z_2'M_{\widehat{X}}Z_2\right)^{-1}Z_2'M_{\widehat{X}}\widehat{u}_1\\ &= \left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widehat{u}_1\\ &= \left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'u,\end{aligned}$$

and hence the OLS estimate $\widehat{\eta}_z^1$ is the same for all $\widehat{\beta}_1$. Specifying the robust variance estimator for $\widehat{\eta}_z^1$ as $V\widehat{a}r_{r,u}\left(\widehat{\eta}_z^1\right)=\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{\Omega}_{21}\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}$, it follows that

$$S_{r,u}\left(\widehat{\eta}_z^1\right)=\widehat{\eta}_z^{1\prime}V\widehat{a}r_{r,u}\left(\widehat{\eta}_z^1\right)\widehat{\eta}_z^1=S_{r,\widehat{x}}\left(\widehat{\beta}_1\right)=J\left(\widehat{\beta}_2\right).$$

Wooldridge (1995) specified the 2SLS robust score test as $S_{r,u}\left(\widehat{\eta}_z^{2sls}\right)$. Baum, Schaffer and Stillman (2007) noted the equivalence of $S_{r,u}\left(\widehat{\eta}_z^{2sls}\right)$ and $J\left(\widehat{\beta}_{gmm}\right)$, but did not provide a proof.

From (3) it follows that if the variance of $\widehat{\eta}_z^{2sls}$ is specified under homoskedasticity as $V\widehat{a}r_u\left(\widehat{\eta}_z^1\right) = \left(\widehat{u}_1'\widehat{u}_1/n\right)\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}$, then

$$S_u\left(\widehat{\eta}_z^1\right) = \widehat{\eta}_z^{1\prime}\left(V\widehat{a}r_u\left(\widehat{\eta}_z^1\right)\right)^{-1}\widehat{\eta}_z^1 = S\left(\widehat{\beta}_1\right).$$

Further versions of the overidentification test can be obtained from (12), with the variance estimator based on the residuals $\widehat{\xi}_1$ instead of $\widehat{u}_1$. The residuals $\widehat{\xi}_1$ are given by

$$\widehat{\xi}_1 = M_Z\widehat{u}_1 = M_Z\left(y - X\widehat{\beta}_1\right) = \widehat{v}_y - \widehat{V}\widehat{\beta}_1,$$

where $\widehat{v}_y$ are the OLS residuals in the first-stage model for y. Under the null that $\pi_y = \Pi\beta$, it follows that $v_{y,i} - v_i'\beta = y_i - x_i'\beta$ and $\widehat{\xi}_1$ is the residual used in Basmann's (1960) statistic. If we specify $V\widehat{a}r_v\left(\widehat{\eta}_{z,}\right) = \left(\widehat{\xi}_1'\widehat{\xi}_1/n\right)\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}$, then

$$\begin{aligned} S_v\left(\widehat{\eta}_z^1\right) &= \widehat{\eta}_z^{1\prime}\left(V\widehat{a}r_v\left(\widehat{\eta}_z^1\right)\right)^{-1}\widehat{\eta}_z^1 \\ &= \frac{\widehat{u}_1'\widetilde{Z}_2\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widehat{u}_1}{\widehat{\xi}_1'\widehat{\xi}_1/n} = S_v\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z-k_x}, \end{aligned}$$

under the maintained assumption of conditional homoskedasticity. $S_v\left(\widehat{\beta}_1\right)$ is Bassman's statistic, where we have refrained from a degrees of freedom correction in the denominator.

The robust version of the test is obtained by specifying

$$V\widehat{a}r_{r,v}\left(\widehat{\eta}_z^1\right) = \left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{\Omega}_2\left(\widehat{\xi}_1\right)\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1},$$

where $\widetilde{\Omega}_2\left(\widehat{\xi}_1\right) = \sum_{i=1}^n \widehat{\xi}_{1i}^2\widetilde{z}_{2i}\widetilde{z}_{2i}'$, and

$$\begin{aligned} S_{r,v}\left(\widehat{\eta}_z^1\right) &= \widehat{\eta}_z^{1\prime}\left(V\widehat{a}r_{r,v}\left(\widehat{\eta}_z^1\right)\right)^{-1}\widehat{\eta}_z^1 \\ &= \widehat{u}_1'\widetilde{Z}_2\left(\widetilde{\Omega}_2\left(\widehat{\xi}_1\right)\right)^{-1}\widetilde{Z}_2'\widehat{u}_1 = S_{r,v}\left(\widehat{\beta}_1\right) \xrightarrow{d} \chi^2_{k_z-k_x}. \end{aligned}$$

From Proposition 1 and the proof in the Appendix, it follows that $S_{r,v}\left(\widehat{\eta}_z^1\right)$ is the same as Hansen's J-test based on the two-step GMM estimator

$$\widehat{\beta}_{2,v} = \arg\min_\beta\,(y - X\beta)'Z\left(\widehat{\Omega}\left(\widehat{\xi}_1\right)\right)^{-1}Z'(y - X\beta),$$

where

$$\widehat{\Omega}\left(\widehat{\xi}_1\right) = \sum_{i=1}^n \widehat{\xi}_{1i}^2 z_i z_i'.$$

Under Assumptions 1 to 5, we have that $\mathrm{r}\begin{pmatrix} \pi_y & \Pi \end{pmatrix} = \mathrm{r}(\Pi) = k_x$, as $\pi_y = \Pi\beta$. For reference in the section concerning the Kleibergen-Paap test below, let

$$\widehat{\eta}^1 = (Z'Z)^{-1} Z'\widehat{u}_1 = \widehat{\pi}_y - \widehat{\Pi}\widehat{\beta}_1.$$

To focus on the overidentying restrictions, let $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$, with Z_1 and $n \times k_x$ matrix. Partition $\widehat{\Pi} = \begin{bmatrix} \widehat{\Pi}_1' & \widehat{\Pi}_2' \end{bmatrix}'$ and $\widehat{\pi}_y$ and $\widehat{\eta}^1$ accordingly. $\widehat{\Pi}_1$ is a $k_x \times k_x$ matrix. Then

$$\widehat{\eta}^1 = \begin{pmatrix} \widehat{\eta}_1^1 \\ \widehat{\eta}_2^1 \end{pmatrix} = \begin{pmatrix} \widehat{\pi}_{y1} - \widehat{\Pi}_1\widehat{\beta}_1 \\ \widehat{\pi}_{y2} - \widehat{\Pi}_2\widehat{\beta}_1 \end{pmatrix}.$$

Let

$$\begin{aligned} Z^* &= \begin{bmatrix} \widehat{X} & Z_2 \end{bmatrix} = ZD \\ D &= \begin{bmatrix} \widehat{\Pi}_1 & 0 \\ \widehat{\Pi}_2 & I_{k_z-k_x} \end{bmatrix}; \; D^{-1} = \begin{bmatrix} \widehat{\Pi}_1^{-1} & 0 \\ -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix}. \end{aligned}$$

Then

$$\begin{aligned} \begin{pmatrix} \widehat{\eta}_x^1 \\ \widehat{\eta}_z^1 \end{pmatrix} &= (Z^{*\prime}Z^*)^{-1} Z^{*\prime}\widehat{u}_1 \\ &= D^{-1}(Z'Z)^{-1} Z'\widehat{u}_1 \\ &= \begin{pmatrix} \widehat{\Pi}_1^{-1}\widehat{\eta}_1^1 \\ \widehat{\eta}_2^1 - \widehat{\Pi}_2\widehat{\Pi}_1^{-1}\widehat{\eta}_1^1 \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} \widehat{\eta}_z^1 &= \begin{bmatrix} -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix} \widehat{\eta}^1 \\ &= \begin{bmatrix} -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix} \widehat{\Pi}_{yx}\widehat{\psi}_1 \end{aligned}$$

where $\widehat{\Pi}_{yx} = \begin{bmatrix} \widehat{\pi}_y & \widehat{\Pi} \end{bmatrix}$ and $\widehat{\psi}_1 = \begin{pmatrix} 1 & -\widehat{\beta}_1 \end{pmatrix}'$.

As

$$\begin{bmatrix} -\widehat{\Pi}_2\widehat{\Pi}_1^{-1} & I_{k_z-k_x} \end{bmatrix} \widehat{\Pi}_{yx} = \begin{bmatrix} \widehat{\pi}_{y2} - \widehat{\Pi}_2\widehat{\Pi}_1^{-1}\widehat{\pi}_{y1} & 0 \end{bmatrix},$$

it follows that

$$\widehat{\eta}_z^1 = \widehat{\pi}_{y2} - \widehat{\Pi}_2\widehat{\Pi}_1^{-1}\widehat{\pi}_{y1},$$

a simple analog estimator for $\gamma = \pi_{y2} - \Pi_2\Pi_1^{-1}\pi_{y1}$ in (5).

3.1 Behaviour of Overidentification Test in Underidentified Models

Staiger and Stock (1997) derived the limiting distribution of the Sargan and Basmann tests $S\left(\widehat{\beta}_{2sls}\right)$ and $S_v\left(\widehat{\beta}_{2sls}\right)$ under the null that $\Pi = 0$, and hence $\mathrm{r}(\Pi) = 0$, and under the local weak instrument asymptotics representation $\Pi = C/\sqrt{n}$, where C is a matrix of constants.

Let z_u a k_z-vector and Z_V a $k_z \times k_x$ matrix of random variables. Let $z_V = \mathrm{vec}(Z_V)$. The random variable $\begin{pmatrix} z_u' & z_V' \end{pmatrix}'$ is distributed as

$$\begin{pmatrix} z_u \\ z_V \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho' \\ \rho & I_{k_x} \end{pmatrix} \otimes I_{k_z}\right). \tag{13}$$

Let

$$b = \left(Z_V' Z_V\right)^{-1} Z_V' z_u.$$

Then, when $\Pi = 0$, and $E(z_i u_i) = 0$, the result is

$$\left\{S\left(\widehat{\beta}_{2sls}\right), S_v\left(\widehat{\beta}_{2sls}\right)\right\} \xrightarrow{d} S^0 = \frac{(z_u - Z_V b)'(z_u - Z_V b)}{1 - 2\rho' b + b' b},$$

see Theorem 3 in Staiger and Stock (1997).

For $k_x = 1$, Figure 1 shows the rejection frequencies when using the 5% critical value of the $\chi^2_{k_z-1}$ distribution when the distribution is that of S^0. The latter is simulated by $30,000$ draws from (13), for values of $\rho = 0, 0.1, 0.2, ..., 0.9$, and $k_z = 3, 10, 20, ..., 50$. As is clear from the figure, the overidentification test under-rejects when the model is underidentified.

4 Underidentification Tests

Tests for underidentification of $H_0 : \mathrm{r}(\Pi) = k_x - 1$ as discussed next are tests for overidentifying restrictions in linear IV models, regressing one of the endogenous explanatory variables on the other endogenous explanatory variables. The instruments used are the full set Z and hence these models are always overidentified, with $k_z - k_x + 1 \geq 1$.

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 $V =$

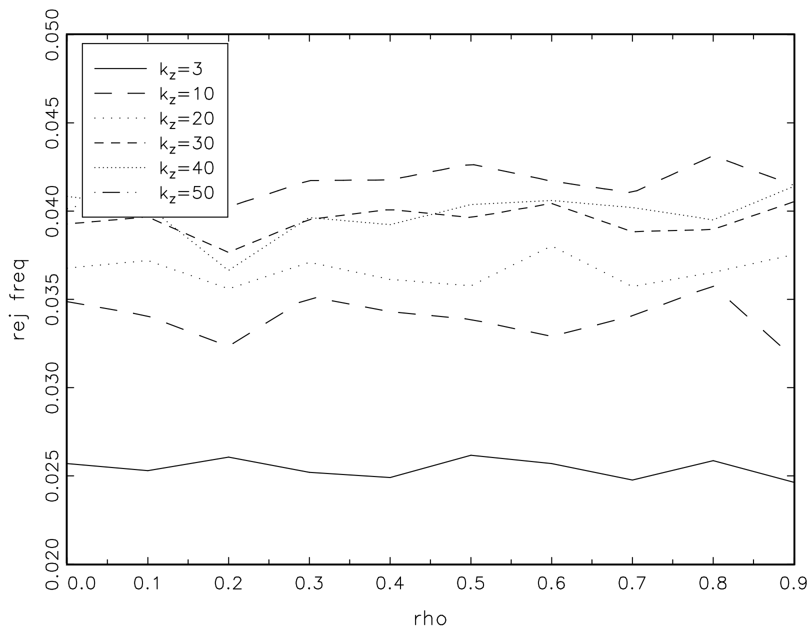

Figure 1: Rejection frequency at 5% when distribution is that of S^0. $k_x = 1$.

$\begin{bmatrix} v_1 & V_2 \end{bmatrix}$ and $\Pi = \begin{bmatrix} \pi_1 & \Pi_2 \end{bmatrix}$. Assuming $\delta_1^* \neq 0$, then $\pi_1 = \Pi_2 \delta$, with $\delta = -\delta_2^*/\delta_1^*$ and hence

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 = Z\Pi_2\delta + v_1 \\ &= X_2\delta + v_1 - V_2\delta \\ &= X_2\delta + \varepsilon_1. \end{aligned} \tag{14}$$

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

$$E(z_i\varepsilon_{1i}) = E(z_i(v_{1i} - v_{2i}'\delta)) = 0, \tag{15}$$

as $E(z_i v_i') = 0$ from the linear projection (2).

Let $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$, with Z_1 a $n \times (k_x - 1)$ matrix and Z_2 a $n \times (k_z - k_x + 1)$ matrix. Further 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

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 = Z_1\pi_{11} + Z_2\pi_{12} + v_1 \\ &= Z_1\Pi_{21}\Pi_{21}^{-1}\pi_{11} + Z_2\pi_{12} + v_1 \\ &= X_2\Pi_{21}^{-1}\pi_{11} + Z_2\left(\pi_{12} - \Pi_{22}\Pi_{21}^{-1}\pi_{11}\right) + v_1 - V_2\Pi_{21}^{-1}\pi_{11} \\ &= X_2\kappa + Z_2\gamma + w_1, \end{aligned} \tag{16}$$

where $\kappa = \Pi_{21}^{-1}\pi_{11}$. When $\pi_1 = \Pi_2\delta$, then $\kappa = \delta$ and $\gamma = 0$. When $\pi_1 \neq \Pi_2\delta$ the overidentifying instruments are in general not "valid" instruments in (14). Note the similarity of model specification (16) and (5) in the previous section, which formed the basis for the score test of overidentifying restrictions.

The intuition of orthogonality condition (15) 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.

4.1 LIML

As we will show below, standard underidentification tests as often reported in applied work are tests for the overidentification restrictions in model (14) after estimation of δ by the Limited Information Maximum Likelihood (LIML) estimator. The LIML estimator is invariant to normalisation, i.e. the test results are invariant to which endogenous regressor x_j has been chosen as the dependent variable in (14). The moment condition $E(z_i\varepsilon_{1i}) = 0$ can then of course be tested using standard tests for overidentification in linear IV models estimated by LIML.

We initially make the assumption of conditional homoskedasticity, $E[v_i v_i' | z_i] = \Sigma_v$. As before, let $\widehat{\Pi} = (Z'Z)^{-1} Z'X$ be the OLS estimator of Π. Denote $\pi = \text{vec}(\Pi)$ and $\widehat{\pi} = \text{vec}\left(\widehat{\Pi}\right)$. Then under the standard assumptions, we have that

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

where $V_\pi = \Sigma_v \otimes Q_{zz}^{-1}$. Denote the first-stage residuals $\widehat{V} = X - Z\widehat{\Pi}$ and let $\widehat{\Sigma}_v = \widehat{V}'\widehat{V}/n$. An estimator for the variance of $\widehat{\pi}$ is therefore $V\widehat{ar}(\widehat{\pi}) = \widehat{\Sigma}_v \otimes (Z'Z)^{-1}$, as $\text{plim}\left(nV\widehat{ar}(\widehat{\pi})\right) = V_\pi$. Partition $\widehat{\Pi}$ as $\widehat{\Pi} = \left[\begin{array}{cc} \widehat{\pi}_1 & \widehat{\Pi}_2 \end{array}\right]$ and let $\pi_2 = \text{vec}(\Pi_2)$ and $\widehat{\pi}_2 = \text{vec}\left(\widehat{\Pi}_2\right)$.

There are various estimating criteria that all result in the same LIML estimator for δ. Following Alonso-Borrego and Arellano (1999), these can be seen to be minimum distance or continuous updating criteria. The first set uses the first-stage residuals $\widehat{V}$.

The minimum distance version of the estimator is given by

$$\begin{aligned}
\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) &= \arg\min_{\delta,\Pi_2} MD_v\left(\delta, \Pi_2\right) \\
MD_v\left(\delta, \Pi_2\right) &= \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix}' \left(V\widehat{a}r\left(\widehat{\pi}\right)\right)^{-1} \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix} \\
&= \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix}' \left(\widehat{\Sigma}_v^{-1} \otimes \left(Z'Z\right)\right) \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix}.
\end{aligned} \tag{17}$$

Concentrating out for the estimator of Π_2 leads to the equivalent continuous updating criterion

$$\begin{aligned}
\widetilde{\delta}_L &= \arg\min_{\delta} S_v\left(\delta\right) \\
S_v\left(\delta\right) &= \frac{\left(x_1 - X_2\delta\right)' P_Z \left(x_1 - X_2\delta\right)}{\left(\widehat{v}_1 - \widehat{V}_2\delta\right)' \left(\widehat{v}_1 - \widehat{V}_2\delta\right)/n}
\end{aligned} \tag{18}$$

Let $\widehat{\Sigma}_x = X'X/n$. The second set is then given by the minimum distance criterion

$$\begin{aligned}
\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) &= \arg\min_{\delta,\Pi_2} MD_x\left(\delta, \Pi_2\right) \\
MD_x\left(\delta, \Pi_2\right) &= \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix}' \left(\widehat{\Sigma}_x^{-1} \otimes \left(Z'Z\right)\right) \begin{pmatrix} \widehat{\pi}_1 - \Pi_2\delta \\ \widehat{\pi}_2 - \pi_2 \end{pmatrix}
\end{aligned} \tag{19}$$

and the continuous updating criterion

$$\begin{aligned}
\widetilde{\delta}_L &= \arg\min_{\delta} S_x\left(\delta\right) \\
S_x\left(\delta\right) &= \frac{\left(x_1 - X_2\delta\right)' P_Z \left(x_1 - X_2\delta\right)}{\left(x_1 - X_2\delta\right)' \left(x_1 - X_2\delta\right)/n}.
\end{aligned} \tag{20}$$

From (17) and (18) we get Basmann's overidentification test

$$MD_v\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) = S_v\left(\widetilde{\delta}_L\right) \xrightarrow{d} \chi_{k_z - k_x + 1},$$

and from (19) and (20) we obtain the Sargan equivalent

$$MD_x\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) = S_x\left(\widetilde{\delta}_L\right) \xrightarrow{d} \chi_{k_z - k_x + 1},$$

under the null that $E\left(z_i\varepsilon_{1i}\right) = 0$, or equivalently that $\mathrm{r}\left(\Pi\right) = k_x - 1$, and under the maintained assumption of homoskedasticity. Again, under the null,

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

The Cragg-Donald (CD) statistic for $H_0 : \mathrm{r}(\Pi) = k_x - 1$, under homoskedasticity, is given by the Basmann test, $MD_v\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) = S_v\left(\widehat{\delta}_L\right)$, see Cragg and Donald (1993). Let δ^* be as defined above, with $\Pi\delta^* = 0$. Then

$$\begin{aligned} \mathrm{CD} &= S_v\left(\widehat{\delta}_L\right) = \min_{\delta^*, \|\delta^*\|=1} \frac{\delta^{*\prime} X' P_Z X \delta^*}{\delta^{*\prime} \widehat{\Sigma}_v \delta^*} \\ &= \min \operatorname{eval}\left(\widehat{\Sigma}_v^{-1/2} X' P_Z X \widehat{\Sigma}_v^{-1/2}\right) \\ &= \min \operatorname{eval}\left(\widehat{\Sigma}_v^{-1/2} \widehat{\Pi}' Z' Z \widehat{\Pi} \widehat{\Sigma}_v^{-1/2}\right), \end{aligned} \tag{21}$$

where $\min \operatorname{eval}(A)$ denotes the minimum eigenvalue of a matrix A. CD is therefore the minimum eigenvalue of the matrix analog of the first-stage F-statistic, see e.g. Stock and Yogo (2005).

Equivalently, we get for $S_x\left(\widehat{\delta}_L\right)$

$$\begin{aligned} S_x\left(\widetilde{\delta}_L\right) &= \min \operatorname{eval}\left(\widehat{\Sigma}_x^{-1/2} X' P_Z X \widehat{\Sigma}_x^{-1/2}\right) \\ &= n \times \min \operatorname{eval}\left(\left(X'X\right)^{-1/2} X' P_Z X \left(X'X\right)^{-1/2}\right). \end{aligned}$$

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 $H_0 : \mathrm{r}(\Pi) = k_x - 1$ 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

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

As $\min \operatorname{eval}\left(\left(X'X\right)^{-1} X' P_Z X\right) = \min \operatorname{eval}\left(\left(X'X\right)^{-1/2} X' P_Z X \left(X'X\right)^{-1/2}\right)$, it follows that

$$S\left(\widetilde{\delta}_L\right) = n\widehat{\rho}^2_{k_x} \xrightarrow{d} \chi^2_{k_z - k_x + 1}. \tag{22}$$

$S\left(\widetilde{\delta}_L\right)$ is thus closely related to the Anderson (1951) likelihood ratio test for $H_0 : \rho_{k_x} = 0$,

$$LR\left(\widehat{\rho}^2_{k_x}\right) = -n \log\left(1 - \widehat{\rho}^2_{k_x}\right)$$

see also Anderson (1984) and the discussion in Hall, Rudebusch and Wilcox (1996). Result (22) modifies the statement in Kleibergen and Paap (2006), as the CD statistic is not the canonical correlation test, as is clear from (21). Note that $V\widehat{ar}_x(\widehat{\pi}) = \widehat{\Sigma}_x \otimes \left(Z'Z\right)^{-1}$

is only a valid estimator of the variance of $\widehat{\pi}$ if $\pi = 0$, which is not the null hypothesis considered here. However, $\pi \neq 0$ does not invalidate the minimum distance/LIML estimator and associated test statistic $MD_x\left(\widetilde{\delta}_L, \widetilde{\Pi}_{2,L}\right) = S_x\left(\widetilde{\delta}_L\right) = n\widehat{\rho}^2_{k_x}$.

The Anderson-Rubin overidentification test is a likelihood ratio test and given by

$$\mathrm{AR}\left(\widetilde{\lambda}\right) = n\log\left(\widetilde{\lambda}\right), \tag{23}$$

where, for model (14), $\widetilde{\lambda} = \min \mathrm{eval}\left(\left(X'M_ZX\right)^{-1/2}X'X\left(X'M_ZX\right)^{-1/2}\right)$ and equal to

$$\widetilde{\lambda} = \frac{\widetilde{\varepsilon}'_{1,L}\widetilde{\varepsilon}_{1,L}}{\widetilde{\varepsilon}'_{1,L}M_Z\widetilde{\varepsilon}_{1,L}}, \tag{24}$$

with $\widetilde{\varepsilon}_{1,L} = x_1 - X_2\widetilde{\delta}_L$, see Anderson and Rubin (1950), and as detailed in Davidson and MacKinnon (1993, pp. 647-648). Under the null, $H_0 : E\left(z_i\varepsilon_{1i}\right) = 0$, $\mathrm{AR}\left(\widetilde{\lambda}\right) \xrightarrow{d} \chi_{k_z-k_x+1}$. $S\left(\widetilde{\delta}_L\right)$ is also closely related to $\mathrm{AR}\left(\widetilde{\lambda}\right)$, as from (24) it follows that $S\left(\widetilde{\delta}_L\right) = n\left(\widetilde{\lambda}-1\right)/\widetilde{\lambda}$ and $\log\left(\widetilde{\lambda}\right) \approx \left(\widetilde{\lambda}-1\right)/\widetilde{\lambda}$, see Baum, Schaffer and Stillman (2007).

Let $\widetilde{X}_{2,L} = Z\widetilde{\Pi}_{2,L}$. As is well-known, $\widetilde{\delta}_L = \left(\widetilde{X}'_{2,L}X_{2,L}\right)^{-1}\widetilde{X}'_{2,L}y$, see e.g. Bowden and Turkington (1984, p.113). It follows that

$$\widetilde{X}'_{2,L}\widetilde{\varepsilon}_{1,L} = 0.$$

In fact, we show in the Appendix that $\widetilde{X}'_{2,L}X_{2,L} = \widetilde{X}'_{2,L}\widetilde{X}_{2,L}$, and hence

$$\widetilde{\delta}_L = \left(\widetilde{X}'_{2,L}\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}'_{2,L}y.$$

Let $Z = \begin{bmatrix} \widetilde{X}_{2,L} & Z_2 \end{bmatrix}$, then

$$S_x\left(\widetilde{\delta}_L\right) = \frac{\widetilde{\varepsilon}'_{1,L}Z_2\left(Z'_2M_{\widetilde{X}_{2,L}}Z_2\right)^{-1}Z'_2\widetilde{\varepsilon}_{1,L}}{\widetilde{\varepsilon}'_{1,L}\widetilde{\varepsilon}_{1,L}/n}; \tag{25}$$

$$S_v\left(\widetilde{\delta}_L\right) = \frac{\widetilde{\varepsilon}'_{1,L}Z_2\left(Z'_2M_{\widetilde{X}_{2,L}}Z_2\right)^{-1}Z'_2\widetilde{\varepsilon}_{1,L}}{\widetilde{\xi}'_{1,L}\widetilde{\xi}_{1,L}/n}, \tag{26}$$

where $\widetilde{\xi}_{1,L} = \widehat{v}_1 - \widehat{V}_2\widetilde{\delta}_L$.

Let $\widetilde{Z}_{2,L} = M_{\widetilde{X}_{2,L}}Z_2$ with rows $\widetilde{z}'_{2L,i}$. Versions of the test statistics, robust to general forms of heteroskedasticity are readily obtained as

$$S_{x,r}\left(\widetilde{\delta}_L\right) = \widetilde{\varepsilon}'_{1,L}Z_2\left(\widetilde{\Omega}_{2,L}\left(\widetilde{\varepsilon}_{1,L}\right)\right)^{-1}Z'_2\widetilde{\varepsilon}_{1,L}; \tag{27}$$

$$S_{v,r}\left(\widetilde{\delta}_L\right) = \widetilde{\varepsilon}'_{1,L}Z_2\left(\widetilde{\Omega}_{2,L}\left(\widetilde{\xi}_{1,L}\right)\right)^{-1}Z'_2\widetilde{\varepsilon}_{1,L}, \tag{28}$$

with $\widetilde{\Omega}_{2,L}\left(\widetilde{\varepsilon}_{1,L}\right)=\sum_{i=1}^{n}\widehat{\varepsilon}_{1L,i}^{2}\widetilde{z}_{2L,i}\widetilde{z}_{2L,i}'$ and $\widetilde{\Omega}_{2,L}\left(\widetilde{\xi}_{1,L}\right)=\sum_{i=1}^{n}\widehat{\xi}_{1L,i}^{2}\widetilde{z}_{2L,i}\widetilde{z}_{2L,i}'$.

As $Z_2'\widetilde{\varepsilon}_{1,L}=Z_2'M_{\widetilde{X}_{2,L}}\widetilde{\varepsilon}_{1,L}$ it follows that we can obtain these four test statistics as in Section 3 from testing $H_0:\eta_z=0$ after the OLS regression of the specification

$$\widetilde{\varepsilon}_{1,L}=\widetilde{X}_{2,L}\eta_x+Z_2\eta_z+\xi_1. \tag{29}$$

For the score test for $H_0:\gamma=0$ in (16), based on the vector of contrasts $Z_2'\widetilde{\varepsilon}_{1,L}$, note that

$$\begin{aligned} Z_2'\widetilde{\varepsilon}_{1,L} &= Z_2'\left(I_n-X_2\left(\widetilde{X}_{2,L}'\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}_{2,L}'\right)\varepsilon_1 \\ &= Z_2'\left(I_n-\widehat{X}_2\left(\widetilde{X}_{2,L}'\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}_{2,L}'\right)\varepsilon_1. \end{aligned}$$

As $Z_2'\widehat{X}_2\neq Z_2'\widetilde{X}_2$, the robust score test alternative version of, for example, $S_{x,r}\left(\widetilde{\delta}_L\right)$ is given by

$$S_{x,r}^{*}\left(\widetilde{\delta}_L\right)=\widetilde{\varepsilon}_{1,L}'Z_2\left(\widetilde{\Omega}_{2,L}^{*}\left(\widetilde{\varepsilon}_{1,L}\right)\right)^{-1}Z_2'\widetilde{\varepsilon}_{1,L},$$

with $\widetilde{\Omega}_{2,L}^{*}\left(\widetilde{\varepsilon}_{1,L}\right)=\sum_{i=1}^{n}\widehat{\varepsilon}_{1L,i}^{2}\widetilde{z}_{2L,i}^{*}\widetilde{z}_{2L,i}^{*\prime}$, with $\widetilde{z}_{2L,i}^{*\prime}$ the rows of

$$\widetilde{Z}_{2,L}^{*}=\left(I_n-\widetilde{X}_{2,L}\left(\widetilde{X}_{2,L}'\widetilde{X}_{2,L}\right)^{-1}\widehat{X}_{2,L}'\right)Z_2.$$

Clearly, as

$$\operatorname{plim}\left(Z_2'\widehat{X}_2/n\right)=\operatorname{plim}\left(Z_2'\widetilde{X}_{2,L}/n\right),$$

$S_{x,r}\left(\widetilde{\delta}_L\right)$ is asymptotically equivalent to $S_{x,r}^{*}\left(\widetilde{\delta}_L\right)$ and both converge in distribution to a $\chi^2_{k_z-k_x-1}$ distributed random variable. $S_{x,r}^{*}\left(\widetilde{\delta}_L\right)$ is the same as the Hansen J-test when using $\widetilde{\delta}_L$ as the one-step estimator, but unlike $S_{x,r}\left(\widetilde{\delta}_L\right)$ it is not invariant to the choice of x_j as the left-hand side variable in (14).

We next show that the robust versions of the Kleibergen-Paap test, as often reported in the literature, are equal to $S_{x,r}\left(\widetilde{\delta}_L\right)$ and $S_{v,r}\left(\widetilde{\delta}_L\right)$ as defined in (27 and (28) respectively.

4.2 Kleibergen-Paap Test

Let G and F be $k_z\times k_z$ and $k_x\times k_x$ finite non-singular matrices respectively, and define

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

For testing $H_0 : \mathrm{r}(\Pi) = q$, Kleibergen and Paap (2006) (KP) propose use of the singular value decomposition (SVD)

$$\Theta = USW',$$

where U and W are $k_z \times k_z$ and $k_x \times k_x$ orthonormal matrices respectively, and S is a $k_z \times k_x$ 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 $q \times k_x$ matrix, $A_{q,\perp}$ is a $k_z \times (k_z - q)$ matrix, Λ_q is $(k_z - q)(k_x - q)$ matrix, $B_{q,\perp}$ is a $(k_x - q) \times k_x$ matrix; $A_q' A_{q,\perp} = 0$ and $B_{q,\perp} B_q' = 0$. As $\Lambda_q = 0$ under the null $H_0 : \mathrm{r}(\Pi) = q$, the KP test is test for $H_0 : \mathrm{vec}(\Lambda_q) = 0$.

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

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

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 estimate of the variance of $\widehat{\pi}$.

For $q = k_x - 1$, we show in Appendix that various versions of the KP test are versions of the score tests as can be obtained from specification (29), generalised as

$$\widetilde{\varepsilon}_1^{GF} = x_1 - X_2 \widetilde{\delta}^{GF} = \widetilde{X}_2^{GF} \eta_x^{GF} + Z_2 \eta_z^{GF} + \xi_1^{GF},$$

where $\widetilde{\Pi}_2^{GF}$ and $\widetilde{\delta}^{GF}$ are given by

$$\left(\widetilde{\Pi}_2^{GF}, \widetilde{\delta}^{GF}\right) = \arg\min_{\Pi_2, \delta} \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \Pi_2 \delta \\ \pi_2 \end{pmatrix} \right)' (F'F \otimes G'G) \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \Pi_2 \delta \\ \pi_2 \end{pmatrix} \right).$$

The estimator $\widetilde{\delta}^{GF}$ can also be obtained as the continuous updating estimator

$$\widetilde{\delta}^{GF} = \arg\min_{\delta} \frac{(x_1 - X_2\delta)' Z (Z'Z)^{-1} G'G (Z'Z)^{-1} Z' (x_1 - X_2\delta)}{\begin{pmatrix} 1 & -\delta' \end{pmatrix}' F'F \begin{pmatrix} 1 & -\delta' \end{pmatrix}'}. \tag{30}$$

It is therefore clear choosing F and G such that $F'F = \widehat{\Sigma}_v^{-1}$ or $F'F = \widehat{\Sigma}_x^{-1}$ and $G'G = Z'Z$ results in the LIML estimators for δ and Π_2. Choosing $F'F = I_{k_x}$ and $G'G = Z'Z$ results in the symmetrically normalised 2SLS estimator, see Alonso-Borrego and Arellano (1999).

The KP tests commonly reported in standard estimation routines, like ivreg2 in Stata, is based on the LIML normalisation and hence are equal to $S_{x,r}\left(\widetilde{\delta}_L\right)$ or $S_{v,r}\left(\widetilde{\delta}_L\right)$.

4.3 CUE

The robust version of the Cragg-Donald statistic is obtained by using the robust variance estimator of $\widehat{\pi}$,

$$V\widehat{a}r_r\left(\widehat{\pi}\right) = \left(I_{k_x} \otimes \left(Z'Z\right)^{-1}\right)\left(\sum_{i=1}^{n}\left(\widehat{v}_i\widehat{v}_i'\right) \otimes \left(z_i z_i'\right)\right)\left(I_{k_x} \otimes \left(Z'Z\right)^{-1}\right),$$

as the weight matrix in the minimum distance criterion (17). As this variance no longer has a Kronecker product form, the resulting estimates for Π_2 and δ cannot be obtained using simple eigenvalue or SVD methods, and will have to be obtained using iterative optimisation techniques. The resulting estimator for δ is a continuous updating estimator, CUE, with $\widehat{v}_{i1} - \widehat{v}_{i2}'\delta$ as the residuals in the weight matrix,

$$\begin{aligned}
\widehat{\delta}_{cue,v} &= \arg\min_{\delta} J_v\left(\delta\right) \\
J_v\left(\delta\right) &= \left(x_1 - X_2\delta\right)' Z\left(\sum_{i=1}^{n}\left(\widehat{v}_{1i} - \widehat{v}_{2i}'\delta\right)^2 z_i z_i'\right)^{-1} Z'\left(x_1 - X_2\delta\right),
\end{aligned}$$

and

$$J_v\left(\widehat{\delta}_{cue,v}\right) \xrightarrow{d} \chi^2_{k_z - k_x + 1}.$$

under the null that $E\left(z_i\varepsilon_{1i}\right) = 0$. As with the LIML estimator, the alternative, and more commonly used CUE estimator is given by

$$\begin{aligned}
\widehat{\delta}_{cue,x} &= \arg\min_{\delta} J_x\left(\delta\right) \\
J_x\left(\delta\right) &= \left(x_1 - X_2\delta\right)' Z\left(\sum_{i=1}^{n}\left(x_{1i} - x_{2i}'\delta\right)^2 z_i z_i'\right)^{-1} Z'\left(x_1 - X_2\delta\right),
\end{aligned}$$

with

$$J_x\left(\widehat{\delta}_{cue,x}\right) \xrightarrow{d} \chi^2_{k_z - k_x + 1}.$$

$J(\delta, x)$ is obtained from the Cragg-Donald minimum distance specification (17) by specifying the variance of $\widehat{\pi}$ under the null that $\pi = 0$,

$$V\widehat{a}r_r(\widehat{\pi}) = \left(I_{k_x} \otimes (Z'Z)^{-1}\right)\left(\sum_{i=1}^{n}(x_i x_i') \otimes (z_i z_i')\right)\left(I_{k_x} \otimes (Z'Z)^{-1}\right),$$

but, clearly, the distribution of $J_x\left(\widehat{\delta}_{cue,x}\right)$ under the null remains valid also when $\pi \neq 0$.

From the results obtained for the two-step GMM Hansen test in Section (3), it follows that the CUE statistics $J_v\left(\widehat{\delta}_{cue,v}\right)$ and $J_x\left(\widehat{\delta}_{cue,x}\right)$ have the same limiting distribution under local alternatives of the form $\gamma = c/\sqrt{n}$ as the robust score test versions $S_{v,r}\left(\widetilde{\delta}_L\right)$ and $S_{x,r}\left(\widetilde{\delta}_L\right)$, see also Al-Sadoon (2017).

4.3.1 Two-Step GMM

Sanderson and Windmeijer (2016) (SW) proposed conditional F-statistics for testing for underidentification or weak instruments for each endogenous variable separately. Their conditional test statistics can be obtained as Basmann tests. Let $\widehat{\delta}_j$ be the 2SLS estimator of δ_j in the model

$$x_j = X_{-j}\delta_j + \varepsilon_j \tag{31}$$

using instruments Z, where for a general matrix A, $A_{-j} = A \backslash \{a_j\}$ where a_j is the j-th column of A. Let $\widehat{\varepsilon}_j = x_j - X_{-j}\widehat{\delta}_j$. They then considered the Basmann tests

$$S_v\left(\widehat{\delta}_j\right) = \frac{\widehat{\varepsilon}_j' Z (Z'Z)^{-1} Z'\widehat{\varepsilon}_j}{\widehat{\xi}_j'\widehat{\xi}_j/n},$$

where $\widehat{\xi}_j = \widehat{v}_j - \widehat{V}_{-j}\widehat{\delta}_j$, and provided the theory for testing for weak instruments based on the F-test version, $F_j = S_v\left(\widehat{\delta}_j\right)/(k_z - k_x + 1)$. The weak instrument asymptotics they considered was that of $\mathrm{r}(\Pi)$ local to a rank reduction of 1, or $\pi_j = \Pi_{-j}\delta_j + p/\sqrt{n}$. They showed that the F_j can provide additional information to that provided by the CD statistic about the nature of the weak instruments problem.

A simple generalisation to robust tests for underidentification is then to simply compute the two-step Hansen J-tests.

5 Dynamic Panel Data Models

I next consider the panel data model

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

for $i = 1, ..., n$, $t = 1, ...T$, where x_{it} is a k_x-vector that 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 all explanatory variables are endogenous, the available moment conditions at period t are given by

$$E\left(x_i^{t-2}\Delta u_{it}\right), \tag{32}$$

where $x_i^{t-2} = \left(\begin{array}{cccc} x_{i1}' & x_{i2}' & \dots & x_{i,t-2}' \end{array}\right)'$. As Bazzi and Clemens (2013) state, a standard test for weak instruments does not currently exist for this framework and they propose the use of the Kleibergen-Paap underidentification test as a diagnostic for weak instruments. Below, we show that KP test in this clustered data setting is no longer equivalent to the CD test under conditional homoskedasticity. The KP test is based on the pooled LIML estimator, which does not take into account the clustering of the data, and is hence not efficient. The CD test is based on an cluster efficient LIML type estimator.

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 moments (32) for individual i can then be expressed as

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

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 reduced form 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. Let Z be the $n(T-2) \times k_z$ matrix $[Z_i]$ and ΔX the $n(T-2) \times k_x$ matrix $[\Delta X_i]$. The OLS (and SUR) estimator for Π is given by

$$\widehat{\Pi} = (Z'Z)^{-1} Z' \Delta X$$

and

$$\begin{aligned} \widehat{\pi} &= \operatorname{vec}\left(\widehat{\Pi}\right) = \left(I_{k_x} \times (Z'Z)^{-1} Z'\right) \operatorname{vec}(\Delta X) \\ &= \left(\widetilde{Z}'\widetilde{Z}\right)^{-1} \widetilde{Z}' \operatorname{vec}(\Delta X) \end{aligned}$$

where $\widetilde{Z} = (I_{k_x} \otimes Z)$ and hence

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

where

$$V_\pi = \left(I_{k_x} \otimes Q_{ZZ}^{-1}\right) \Omega_{Zv} \left(I_{k_x} \otimes Q_{ZZ}^{-1}\right)$$

where $Q_{ZZ} = E(Z_i'Z_i)$ and

$$\Omega_{Zv} = E\left[\operatorname{vec}(Z_i'V_i) \operatorname{vec}(Z_i'V_i)'\right].$$

Under conditional homoskedasticity,

$$E\left[\operatorname{vec}(V_i) \operatorname{vec}(V_i)' \,|Z_i\right] = E\left[\operatorname{vec}(V_i) \operatorname{vec}(V_i)'\right] = \Sigma_{\operatorname{vec}(V)},$$

where $\Sigma_{\operatorname{vec}(V)}$ is a $(T-2)\,k_x \times (T-2)\,k_x$ matrix, we have that

$$\Omega_{Zv} = E\left[\left(I_{k_x} \otimes Z_i'\right) \Sigma_{\operatorname{vec}(V)} \left(I_{k_x} \otimes Z_i\right)\right].$$

Let

$$\widehat{\Sigma}_{\operatorname{vec}(V)} = \frac{1}{n} \sum_{i=1}^{n} \operatorname{vec}\left(\widehat{V}_i\right) \operatorname{vec}\left(\widehat{V}_i\right)',$$

then

$$V\widehat{ar}(\widehat{\pi}) = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1} \left(\sum_{i=1}^{n} \widetilde{Z}_i' \widehat{\Sigma}_{\operatorname{vec}(V)} \widetilde{Z}_i\right) \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}, \tag{33}$$

where $\widetilde{Z}_i = (I_{k_x} \otimes Z_i)$, is the variance estimator for $\widehat{\pi}$ under conditional homoskedasticity, and

$$V\widehat{ar}_r(\widehat{\pi}) = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\left(\sum_{i=1}^{n}\widetilde{Z}_i' \operatorname{vec}\left(\widehat{V}_i\right)\operatorname{vec}\left(\widehat{V}_i\right)'\widetilde{Z}_i\right)\left(\widetilde{Z}'\widetilde{Z}\right)^{-1} \tag{34}$$

is a robust variance estimator.

Partitioning $\Delta X_i = \begin{bmatrix} \Delta x_{1i} & \Delta X_{2i} \end{bmatrix}$ and, as before, $\Pi = \begin{bmatrix} \pi_1 & \Pi_2 \end{bmatrix}$, then the "LIML" equivalent for the CD statistic in this clustered panel data setting is (17) using variance estimate (33). A robust version of this LIML type CD statistic is again obtained as follows. Let $\widehat{\delta}$ and $\widehat{\Pi}_2$ be the estimators of δ and Π_2 minimising the minimum distance CD criterion. Define the residual

$$\widehat{\varepsilon}_{it} = \Delta x_{1,it} - \Delta x_{2,it}'\widehat{\delta}.$$

Then estimate the parameters in the model

$$\widehat{\varepsilon}_{it} = \left(z_{it}'\widehat{\Pi}_2\right)\eta_x + z_{2,it}'\eta_z + \xi_{it}$$

and construct the robust test based on $\widehat{\eta}_z$ and a cluster robust variance estimate.

Note that the Kleibergen-Paap test is a robust version of the test but based on the "pooled" LIML estimator, i.e. the LIML estimator that does not take into account the within individual correlations.

The robust version of the CD statistic is obtained using robust variance estimate (34). Again this is asymptotically equivalent to Hansen's J-test after estimating the model

$$\Delta x_{1,it} = \Delta x_{2,it}'\delta + \varepsilon_{it}$$

by the CUE, using the same z_{it} instruments as in the original model for Δy_{it}. Under the null that the model is underidentified, $\pi_1 = \Pi_2\delta$, we have that

$$J_v\left(\widehat{\delta}_{CUE,v}\right) \xrightarrow{d} \chi^2_{k_z-k_x+1}.$$

The simple extension of the Sanderson-Windmeijer approach is then to estimate the models

$$\Delta x_{j,it} = \Delta x_{-j,it}'\delta_j + \varepsilon_{it}$$

by two-step GMM, again using the same z_{it} as instruments, and to test the null $H_0 : \pi_j = \Pi_{-j}\delta_j$ with the Hansen J-test. The latter 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. For example, if the original model command is

xtabond2 y l.y x1 x2 i.year, gmm(y x1 x2, lag (2 4)) iv(i.year) nol rob

then the test for $H_0 : \pi_{l.y} = \Pi_{x_1 x_2} \delta_{l.y}$ is obtained as the Hansen test from the GMM estimation

xtabond2 l.y x1 x2 i.year, gmm(y x1 x2, lag(2 4)) iv(i.year) nol rob.

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

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*, http://dx.doi.org/10.1016/j.jeconom.2017.03.002

Andrews, I., 2015, Robust two-step confidence sets, and the trouble with the first stage F-statistic, mimeo, Harvard.

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.

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.

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.

Blundell, R. and S. Bond, 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics 87*, 115-143.

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

Bun, M. 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.

Cragg, J.G. and S.G. Donald, 1993. Testing Identifiability and Specification in Instrumental Variables Models, *Econometric Theory* 9, 222-240.

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

Healy, M.J.R., 1968. Multiple regression with a singular matrix, *Journal of the Royal Statistical Society, Series C* 17, 110-117.

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

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.

Skeels, C.L. and F. Windmeijer, 2016, On the Stock-Yogo Tables, Dept. of Economics Discussion Paper 16/679, University of Bristol.

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.

6 Appendix

6.1 Proof of Proposition 1

Consider the one-step GMM estimator with weight matrix W_n, given by

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

and

$$\widehat{u}_1 = y - X\widehat{\beta}_1.$$

Let

$$\widehat{X}_w = ZW_n^{-1}Z'X,$$

then

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

and

$$\widehat{X}_w'\widehat{u}_1 = \widehat{X}_w y - \widehat{X}_w'X\left(\widehat{X}_w'X\right)^{-1}\widehat{X}_w'y = 0.$$

Let $Z = \left[\begin{array}{cc} \widehat{X}_w & Z_2 \end{array}\right]$, and partition

$$\widehat{\Omega}_{\widehat{\beta}_1} = \sum_{i=1}^{n} \widehat{u}_{1i}^2 z_i z_i'$$

commensurate as

$$\widehat{\Omega}_{\widehat{\beta}_1} = \left[\begin{array}{cc} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{array}\right],$$

where for example $A_{22} = \sum_{i=1}^{n} \widehat{u}_{1i}^2 z_{2i} z_{2i}'$. From standard results for the inverse of a symmetric partitioned matrix, we get for the inverse of $\widehat{\Omega}_{\widehat{\beta}_1}$

$$\widehat{\Omega}_{\widehat{\beta}_1}^{-1} = \left[\begin{array}{cc} B_{11} & B_{12} \\ B_{12}' & B_{22} \end{array}\right],$$

with

$$\begin{aligned} B_{11} &= \left(A_{11} - A_{12}A_{22}^{-1}A_{12}'\right)^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}B_{22}A_{12}'A_{11}^{-1} \\ B_{22} &= \left(A_{22} - A_{12}'A_{11}^{-1}A_{12}\right)^{-1} = A_{22}^{-1} + A_{22}^{-1}A_{12}'B_{11}A_{12}A_{22}^{-1} \\ B_{12} &= -A_{11}^{-1}A_{12}B_{22}, \end{aligned}$$

see e.g. Magnus and Neudecker (2007, p. 12).

First, consider

$$S_r\left(\widehat{\beta}_1\right) = \widehat{u}_1' Z_2 \widetilde{\Omega}_{2w}^{-1} Z_2' \widehat{u}_1,$$

with

$$\widetilde{\Omega}_{2w} = \widetilde{Z}_{2w}' \widehat{H} \widetilde{Z}_{2w},$$

where $\widetilde{Z}_{2w} = \left(I_{k_z} - \widehat{X}_w \left(\widehat{X}_w' X\right)^{-1} X'\right) Z_2 = \left(I_{k_z} - \widehat{X}_w \left(\widehat{X}_w' \widehat{X}\right)^{-1} \widehat{X}'\right) Z_2$ and $\widehat{H}$ the $n \times n$ diagonal matrix $\widehat{H} = \mathrm{diag}\left(\widehat{u}_{1i}^2\right)$. For brevity, define

$$C = \widehat{X}_w' \widehat{X}; \quad D = Z_2' \widehat{X},$$

and note that C is symmetric. Then

$$\widetilde{\Omega}_{2w} = A_{22} - DC^{-1}A_{12} - A_{12}'C^{-1}D' + DC^{-1}A_{11}C^{-1}D'$$

Next, consider Hansen's J-test as given in (7), which can be rewritten as

$$\begin{aligned}
J\left(\widehat{\beta}_2\right) &= y'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'y - y'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1}X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'y \\
&= \widehat{u}_1'Z\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1} - \widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1}X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)Z'\widehat{u}_1.
\end{aligned}$$

As

$$Z'\widehat{u}_1 = \begin{bmatrix} 0 \\ Z_2'\widehat{u}_1 \end{bmatrix},$$

we need to determine $\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)_{22} - \left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1}X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)_{22}$, using the same partitioning as above.

For the first term, we have that $\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)_{22} = B_{22}$. For the second term we have first, noting that $Z'X = \begin{bmatrix} \widehat{X}_w'X \\ Z_2'X \end{bmatrix} = \begin{bmatrix} \widehat{X}_w'\widehat{X} \\ Z_2'\widehat{X} \end{bmatrix} = \begin{bmatrix} C \\ D \end{bmatrix}$,

$$\begin{aligned}
X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X &= CB_{11}C + CB_{12}D + D'B_{12}'C + D'B_{22}D \\
&= CA_{11}^{-1}C + CA_{11}^{-1}A_{12}B_{22}A_{12}'A_{11}^{-1}C - CA_{11}^{-1}A_{12}B_{22}D \\
&\quad -D'B_{22}A_{12}'A_{11}^{-1}C + D'B_{22}D \\
&= CA_{11}^{-1}C + \left(D - A_{12}'A_{11}^{-1}C\right)'B_{22}\left(D - A_{12}'A_{11}^{-1}C\right).
\end{aligned}$$

Using the fact that for suitably defined matrices,

$$\left(E + F'GF\right)^{-1} = E^{-1} - E^{-1}F'\left(G^{-1} + FE^{-1}F'\right)^{-1}FE^{-1},$$

it follows that

$$\begin{aligned}\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1} &= C^{-1}A_{11}C^{-1}\\ &\quad -C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'\\ &\quad \times\left(B_{22}^{-1}+\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'\right)^{-1}\\ &\quad \times\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}.\end{aligned}$$

Note that

$$\begin{aligned}& B_{22}^{-1}+\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'\\ =\;& A_{22}-A_{12}'A_{11}^{-1}A_{12}-DC^{-1}A_{12}-A_{12}'C^{-1}D'+DC^{-1}A_{11}C^{-1}D'+A_{12}'A_{11}^{-1}A_{12}\\ =\;& \widetilde{\Omega}_{2w}.\end{aligned}$$

Next,

$$X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}=\left[\begin{array}{cc} CB_{11}+D'B_{12}' & CB_{12}+D'B_{22}\end{array}\right].$$

As

$$CB_{12}+D'B_{22}=\left(D-A_{12}'A_{11}^{-1}C\right)'B_{22},$$

$\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1}X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)_{22}$ is given by

$$\begin{aligned}& B_{22}\left(D-A_{12}'A_{11}^{-1}C\right)\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'B_{22}\\ =\;& B_{22}\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'B_{22}\\ & -B_{22}\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'\\ & \times\widetilde{\Omega}_{2w}^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)C^{-1}A_{11}C^{-1}\left(D-A_{12}'A_{11}^{-1}C\right)'B_{22}\\ =\;& B_{22}\left(\widetilde{\Omega}_{2w}-B_{22}^{-1}\right)B_{22}-B_{22}\left(\widetilde{\Omega}_{2w}-B_{22}^{-1}\right)\widetilde{\Omega}_{2w}^{-1}\left(\widetilde{\Omega}_{2w}-B_{22}^{-1}\right)B_{22}\\ =\;& B_{22}\widetilde{\Omega}_{2w}B_{22}-B_{22}-B_{22}\widetilde{\Omega}_{2w}B_{22}+B_{22}+B_{22}-\widetilde{\Omega}_{2w}^{-1}\\ =\;& B_{22}-\widetilde{\Omega}_{2w}^{-1}.\end{aligned}$$

Hence,

$$
\begin{aligned}
J\left(\widehat{\beta}_2\right) &= \widehat{u}_1' Z\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1} - \widehat{\Omega}_{\widehat{\beta}_1}^{-1} Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1} X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right) Z'\widehat{u}_1 \\
&= \widehat{u}_1' Z_2\left(\widehat{\Omega}_{\widehat{\beta}_1}^{-1} - \widehat{\Omega}_{\widehat{\beta}_1}^{-1} Z'X\left(X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}Z'X\right)^{-1} X'Z\widehat{\Omega}_{\widehat{\beta}_1}^{-1}\right)_{22} Z_2'\widehat{u}_1 \\
&= \widehat{u}_1' Z_2\left(B_{22} - \left(B_{22} - \widetilde{\Omega}_{2w}^{-1}\right)\right) Z_2'\widehat{u}_1 \\
&= \widehat{u}_1' Z_2 \widetilde{\Omega}_{2w}^{-1} Z_2' \widehat{u}_1 \\
&= S_r\left(\widehat{\beta}_1\right).
\end{aligned}
$$

6.2 Proof of Proposition 2

Let

$$A_w = \widehat{X}_w\left(\widehat{X}_w'\widehat{X}\right)^{-1}\widehat{X}'.$$

Then

$$
\begin{aligned}
\widetilde{Z}_{2w} &= \left(I_n - A_w\right) Z_2 \\
\widetilde{Z}_2 &= M_{\widehat{X}} Z_2 = \left(I_n - P_{\widehat{X}}\right) Z_2.
\end{aligned}
$$

As $A_w A_w = A_w$, $(I_n - A_w)$ is an idempotent, but not symmetric, projection matrix, with $P_{\widehat{X}} A_w = P_{\widehat{X}}$ and $A_w P_{\widehat{X}} = A_w$. Hence

$$
\begin{aligned}
M_{\widehat{X}}\left(I_n - A_w\right) &= \left(I_n - A_w\right), \\
\left(I_n - A_w\right) M_{\widehat{X}} &= M_{\widehat{X}}
\end{aligned}
$$

Therefore, it follows that

$$\widetilde{Z}_{2w} = P_{\widetilde{Z}_2}\widetilde{Z}_{2w} = \widetilde{Z}_2\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widetilde{Z}_{2w}. \tag{35}$$

and $P_{\widetilde{Z}_{2w}}\widetilde{Z}_2 = \widetilde{Z}_2$.

Let

$$
\begin{aligned}
S_r\left(\widehat{\beta}_1\right) &= \widehat{u}_1' Z_2 \widetilde{\Omega}_{2w}^{-1} Z_2' \widehat{u}_1 \\
&= \widehat{u}_1' \widetilde{Z}_{2w}\left(\widetilde{Z}_{2w}'\widehat{H}\widetilde{Z}_{2w}\right)^{-1}\widetilde{Z}_{2w}'\widehat{u}_1,
\end{aligned}
$$

with $\widehat{H} = \operatorname{diag}\left(\widehat{u}_{1i}^2\right)$. From (35) it then follows that

$$\begin{aligned}
S_r\left(\widehat{\beta}_1\right) &= \widehat{u}_1'\widetilde{Z}_{2w}\left(\widetilde{Z}_{2w}'\widehat{H}\widetilde{Z}_{2w}\right)^{-1}\widetilde{Z}_{2w}'\widehat{u}_1 \\
&= \widehat{u}_1'P_{\widetilde{Z}_2}\widetilde{Z}_{2w}\left(\widetilde{Z}_{2w}'P_{\widetilde{Z}_2}\widehat{H}\widetilde{Z}_{2w}P_{\widetilde{Z}_2}\right)^{-1}\widetilde{Z}_{2w}'P_{\widetilde{Z}_2}\widehat{u}_1 \\
&= \widehat{u}_1'\widetilde{Z}_2\left(\widetilde{Z}_2'\widehat{H}\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widehat{u}_1 \\
&= S_{r,\widehat{x}}\left(\widehat{\beta}_1\right),
\end{aligned}$$

as

$$\left(\widetilde{Z}_{2w}'P_{\widetilde{Z}_2}\widehat{H}\widetilde{Z}_{2w}P_{\widetilde{Z}_2}\right)^{-1} = \left(\widetilde{Z}_2'\widetilde{Z}_{2w}\right)^{-1}\widetilde{Z}_2'\widetilde{Z}_2\left(\widetilde{Z}_2'\widehat{H}\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widetilde{Z}_2\left(\widetilde{Z}_{2w}'\widetilde{Z}_2\right)^{-1}.$$

6.3 Proof that $\widetilde{X}_{2,L}'X_{2,L} = \widetilde{X}_{2,L}'\widetilde{X}_{2,L}$

As $\Pi_2\delta = (\delta' \otimes I_{k_z})\operatorname{vec}(\Pi_2) = (\delta' \otimes I_{k_z})\pi_2$, the minimum distance estimator is determined as,

$$\left(\widetilde{\delta}_L, \widetilde{\pi}_{2,L}\right) = \arg\min_{\delta,\pi_2}\left(\widehat{\pi} - \left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]' \otimes I_{k_z}\right)\pi_2\right)'\left(\widehat{\Sigma}^{-1} \otimes Z'Z\right)\left(\widehat{\pi} - \left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]' \otimes I_{k_z}\right)\pi_2\right),$$

where $d = k_x - 1$. The first-order condition for π_2 is then given by

$$\left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]' \otimes I_{k_z}\right)'\left(\widehat{\Sigma}^{-1} \otimes Z'Z\right)\left(\widehat{\pi} - \left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]' \otimes I_{k_z}\right)\pi_2\right) = 0$$

or

$$\widetilde{\pi}_{2,L}(\delta) = \left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]\widehat{\Sigma}^{-1}\left[\begin{array}{cc}\delta & I_d\end{array}\right]' \otimes Z'Z\right)^{-1}\left(\left[\begin{array}{cc}\delta & I_d\end{array}\right]\widehat{\Sigma}^{-1} \otimes Z'Z\right)\widehat{\pi}$$

reulting in

$$\widetilde{\Pi}_{2,L} = \widehat{\Pi}\widehat{\Sigma}^{-1}\left[\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right]'\left(\left[\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right]\widehat{\Sigma}^{-1}\left[\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right]'\right)^{-1}.$$

From the results in Bowden and Turkington (1984, p. 113), we know that

$$\widetilde{\delta}_L = \left(\widetilde{X}_{2,L}'X_{2,L}\right)^{-1}\widetilde{X}_{2,L}'x_1,$$

and hence

$$\left[\begin{array}{cc}\widetilde{\delta}_L & I_{k_2}\end{array}\right] = \left(\widetilde{X}_{2,L}'X_2\right)^{-1}\widetilde{X}_{2,L}'X.$$

Therefore,

$$\begin{aligned}
\widetilde{\Pi}_{2,L} &= \widehat{\Pi}\widehat{\Sigma}^{-1}\left(\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right)'\left(\left(\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right)\widehat{\Sigma}^{-1}\left(\begin{array}{cc}\widetilde{\delta}_L & I_d\end{array}\right)'\right)^{-1} \\
&= \widehat{\Pi}\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\left(X_2'\widetilde{X}_{2,L}\right)^{-1}\left(\left(\widetilde{X}_{2,L}'X_2\right)^{-1}\widetilde{X}_{2,L}'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\left(X_2'\widetilde{X}_{2,L}\right)^{-1}\right)^{-1} \\
&= \widehat{\Pi}\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\left(\widetilde{X}_{2,L}'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}_{2,L}'X_2 \\
&= \left(Z'Z\right)^{-1}Z'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\left(\widetilde{X}_{2,L}'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}_{2,L}'X_2.
\end{aligned}$$

Then the result follows, as

$$\begin{aligned}
\widetilde{X}_{2,L}'\widetilde{X}_{2,L} &= \widetilde{\Pi}_{2,L}'Z'Z\widetilde{\Pi}_{2,L} \\
&= \widetilde{\Pi}_{2,L}'Z'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\left(\widetilde{X}_{2,L}'X\widehat{\Sigma}^{-1}X'\widetilde{X}_{2,L}\right)^{-1}\widetilde{X}_{2,L}'X_2 \\
&= \widetilde{X}_{2,L}'X_2.
\end{aligned}$$

6.4 Kleibergen-Paap Test

It is illustrative to first set $G = I_{k_z}$ and $F = I_{k_x}$, hence $\Theta = \Pi$ and $\widehat{\Theta} = \widehat{\Pi}$. Order the columns of Π such that $\Pi = \left[\begin{array}{cc}\Pi_2 & \pi_1\end{array}\right]$, and likewise for $\widehat{\Pi}$. It then follows from the discussion in KP for the IV model on pages 101 and 102, that for $q = k_x - 1$, and $\pi_1 = \Pi_2\delta$, and let again $d = k_z - k_x + 1$,

$$\begin{aligned}
A_q &= \Pi_2; B_q = \left[\begin{array}{cc}I_{k_x-1} & \delta\end{array}\right] \\
A_qB_q &= \left[\begin{array}{cc}\Pi_2 & \Pi_2\delta\end{array}\right] \\
A_{q,\perp} &= \left(\begin{array}{c}-\left(\Pi_{21}'\right)^{-1}\Pi_{22}' \\ I_d\end{array}\right)\left(I_d + \Pi_{22}\Pi_{21}^{-1}\left(\Pi_{21}'\right)^{-1}\Pi_{22}'\right)^{-1/2} \\
B_{q,\perp} &= \left(\begin{array}{cc}1 & -\delta'\end{array}\right)/\sqrt{1+\delta'\delta} = \psi'/\sqrt{\psi'\psi}.
\end{aligned}$$

For the test statistic, we can ignore the standardisation terms $\left(I_d + \Pi_{22}\Pi_{21}^{-1}\left(\Pi_{21}'\right)^{-1}\Pi_{22}'\right)^{-1/2}$ and $\left(\psi'\psi\right)^{-1/2}$ and it follows that the test statistic is based on

$$\widehat{\Lambda}_q = \left[\begin{array}{cc}-\widetilde{\Pi}_{22}\widetilde{\Pi}_{21}^{-1} & I_d\end{array}\right]\widehat{\Pi}\widetilde{\psi} \tag{36}$$

where the estimators $\widetilde{\Pi}_2$ and $\widetilde{\Pi}_2\widetilde{\delta}$ are determined from

$$\widehat{A}_q\widehat{B}_q = \left[\begin{array}{cc}\widetilde{\Pi}_2 & \widetilde{\Pi}_2\widetilde{\delta}\end{array}\right].$$

We therefore see that $\widehat{\Lambda}_q$ is identical to $\widehat{\eta}_z$ in (29), given the estimates for Π_2 and δ. For this case where $G = I_{k_z}$ and $F = I_{k_x}$, $\widetilde{\Pi}_2$ and $\widetilde{\delta}$ are given by

$$\left(\widetilde{\Pi}_2, \widetilde{\delta}\right) = \arg\min_{\delta,\Pi_2} \left(\left(\begin{array}{c} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{array}\right) - \left(\begin{array}{c} \Pi_2\delta \\ \pi_2 \end{array}\right)\right)' \left(\left(\begin{array}{c} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{array}\right) - \left(\begin{array}{c} \Pi_2\delta \\ \pi_2 \end{array}\right)\right).$$

Exactly the same formula for $\widehat{\Lambda}_q$ as in (36) is obtained for general choices of F and G, only the estimators $\widetilde{\Pi}_2$ and $\widetilde{\delta}$ vary with F and G. Denote these estimators $\widetilde{\Pi}_2^{GF}$ and $\widetilde{\delta}^{GF}$. 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

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

from which it follows that

$$\left(G'G\right)^{-1} G' \left(\widehat{A}_q\widehat{B}_q\right) F \left(F'F\right)^{-1} = \left[\begin{array}{cc} \widetilde{\Pi}_2^{GF} & \widetilde{\Pi}_2^{GF}\widetilde{\delta}^{GF} \end{array}\right].$$

Hence $\mathrm{rk}\left(k_x - 1\right)$ is identical to $S_{r,v}\left(\widetilde{\delta}^{GF}\right)$ when $\widehat{\Omega}_{k_x-1}$ is based on the robust variance estimator of $\widehat{\pi}$.

6.5 Result for CUE

Let

$$V\widehat{ar}_r\left(\widehat{\pi}\right) = \left(I_{k_x} \otimes \left(Z'Z\right)^{-1}\right)\left(\sum_{i=1}^{n} \left(\widehat{v}_i\widehat{v}_i'\right) \otimes \left(z_iz_i'\right)\right)\left(I_{k_x} \otimes \left(Z'Z\right)^{-1}\right),$$

Following Kleibergen and Mavroeidis (2009),

$$\begin{aligned}
CD_r &= \min_{\Pi_2,\delta} \operatorname{vec}\left(\widehat{\Pi} - \Pi_2\left(\begin{array}{cc} \delta & I_{k_x-1} \end{array}\right)\right)' \left(V\widehat{ar}_r\left(\widehat{\pi}\right)\right)^{-1} \operatorname{vec}\left(\widehat{\Pi} - \Pi_2\left(\begin{array}{cc} \delta & I_{k_x-1} \end{array}\right)\right) \\
&= \min_{\Pi_2,\delta} \operatorname{vec}\left(\left(\widehat{\Pi} - \Pi_2\left(\begin{array}{cc} \delta & I_{k_x-1} \end{array}\right)\right)\left(\begin{array}{cc} 1 & 0 \\ -\delta & I_{k_x-1} \end{array}\right)\right)' \\
&\quad \left(\left(\left(\begin{array}{cc} 1 & 0 \\ -\delta & I_{k_x-1} \end{array}\right) \otimes I_{k_z}\right)' V\widehat{ar}_r\left(\widehat{\pi}\right)\left(\left(\begin{array}{cc} 1 & 0 \\ -\delta & I_{k_x-1} \end{array}\right) \otimes I_{k_z}\right)\right)^{-1} \\
&\quad \operatorname{vec}\left(\left(\widehat{\Pi} - \Pi_2\left(\begin{array}{cc} \delta & I_{k_x-1} \end{array}\right)\right)\left(\begin{array}{cc} 1 & 0 \\ -\delta & I_{k_x-1} \end{array}\right)\right)
\end{aligned}$$

resulting in, $\psi = \begin{pmatrix} 1 & -\delta' \end{pmatrix}'$,

$$\begin{aligned} CD_r &= \min_{\delta} \left(\psi'\widehat{\Pi}\right) \left((\psi \otimes I_{k_z})' V\widehat{a}r_r\left(\widehat{\pi}\right)(\psi \otimes I_{k_z})\right)^{-1} \widehat{\Pi}\psi \\ &= \min_{\delta} \left(x_1 - X_2\delta\right)' Z \left(\sum_{i=1}^{n} \left(\widehat{v}_i'\psi\right)^2 z_i z_i'\right)^{-1} Z' \left(x_1 - X_2\delta\right) \\ &= \min_{\delta} \left(x_1 - X_2\delta\right)' Z \left(\sum_{i=1}^{n} \left(\widehat{v}_{i1} - \widehat{v}_{i2}'\delta\right)^2 z_i z_i'\right)^{-1} Z' \left(x_1 - X_2\delta\right). \end{aligned}$$