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

Frank Windmeijer*
Department of Statistics and Nuffield College
University of Oxford
24-29 St Giles', Oxford OX1 3LB, UK
and Cemmap, UCL/IFS, London, UK

March 19, 2021

Abstract

This paper develops the links between overidentification tests, underidentification tests, score tests and the Cragg and Donald (1993, 1997) and Kleibergen and Paap (2006) rank tests in linear instrumental variable (IV) models. For the structural linear model $y = X\beta + u$, with the endogenous explanatory variables partitioned as $X = [x_1 \, X_2]$, this general framework shows that standard underidentification tests are tests for overidentification in an auxiliary linear model, $x_1 = X_2\delta + \varepsilon$, 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 tests for the rank of general parameter matrices. Invariant rank tests are based on the LIML or continuously updated GMM estimators of both structural and first-stage parameters. This insight leads to the proposal of new two-step invariant asymptotically efficient GMM estimators, and a new iterated GMM estimator that, if it converges, 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

*E-mail address: frank.windmeijer@stats.ox.ac.uk.

1 Introduction

It is common practice when reporting estimation results of standard linear instrumental variable (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. For the structural linear model specification $y = X\beta + u$, they are tests on the rank of Π in the linear projection model for the endogenous explanatory variables X on the instruments Z, $X = Z\Pi + V$, with the standard tests testing the null, $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$, where $\mathrm{r}(\Pi)$ denotes the rank of the matrix Π and k_x is the number of explanatory variables.

The main contribution of this paper is as follows. Partition $X = [x_1\, X_2]$, then it is shown in Section 2 that these underidentification tests are tests for overidentification in the auxiliary model $x_1 = X_2\delta + \varepsilon$, estimated by IV methods using Z as instruments. The non-robust Cragg-Donald statistic is then equal to the Sargan (1958) or Basmann (1960) test statistics for overidentifying restrictions after estimating the parameters in the auxiliary model by limited information maximum likelihood (LIML). The robust Cragg-Donald statistic is the Hansen J-statistic (Hansen, 1982), based on the continuously updated generalised method of moments (CU-GMM) estimator. Non-robust refers throughout to the limiting χ^2 distribution of the test statistic under the null being valid under conditional homoskedasticity and no serial correlation only, whereas robust refers to it being valid under conditional heteroskedasticity and/or serial correlation/clustering as well. These LIML and CU-GMM based test statistics are invariant to the choice of normalisation, i.e. the choice of x_j as the dependent variable in the auxiliary regression. Following the results of Newey and West (1987), these tests are score tests for $H_0 : \gamma = 0$ in the specification $x_1 = X_2\delta + Z_o\gamma + \varepsilon$, where Z_o is any $k_z - k_x + 1$ subset of instruments, and it is shown in Section 2.2 that the standard robust Kleibergen-Paap test statistic is a LIML based invariant robust score test statistic.

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 test statistics are equal to the 2SLS based non-robust Sargan or Basmann test statistics and robust score test statistics are equal to two-step

GMM Hansen J-statistics. For the homoskedastic case, the non-robust Cragg-Donald and Sanderson-Windmeijer tests can be used for testing for weak instruments. However, the Stock and Yogo (2005) weak-instrument test results do not apply when using robust test statistics for dealing with non-homoskedasticity, see Bun and de Haan (2010), Andrews (2018) and Kim (2019).

The standard robust Kleibergen-Paap and robust Cragg-Donald score tests achieve invariance by incorporating the LIML or CU-GMM estimator for Π_2 to form the optimal combination of instruments in the auxiliary model, with $X_2 = Z\Pi_2 + V_2$. This differs from the standard two-step GMM framework which uses the OLS estimator for Π_2 in both one-step and two-step estimators for δ. We use this observation to propose in Section 2.2 a two-step invariant asymptotically efficient GMM estimator that is based on the LIML estimator for δ as the one-step estimator and uses the LIML estimator for Π_2 to construct the optimal instruments for the second step. The J-statistic evaluated at this two-step estimator is then an invariant alternative to the standard robust Kleibergen-Paap statistic. Alternatively, as shown in Appendix A, one can update the estimator for Π_2 from the robust first-order conditions of the CU-GMM estimator, which leads to a different invariant two-step estimator and an iterated GMM estimator that, if it converges, converges to the CU-GMM estimator.

Section 2.4 applies the various robust tests for underidentification to dynamic panel data models. 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. Following Holtz-Eakin, Newey, and Rosen (1988), commonly used estimation methods are the Arellano and Bond (1991) and Blundell and Bond (1998) GMM procedures. These can suffer from weak instrument problems 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 (2019). However, underidentification tests have not previously been developed for these models. Bazzi and Clemens (2013) did consider the Kleibergen-Paap test for this setting, but only in a per period cross-sectional setting, which does not result in a valid test for underidentification as it does not take clustering into account. Section 2.4 provides results for an application taken from Acemoglu, Johnson, Robinson, and Yared (2008) and for a Monte Carlo analysis with a design similar to the application.[1]

[1]The results as presented in the applications can be obtained with the Stata modules UNDERID (Schaffer and Windmeijer, 2020), and RANKTEST (Kleibergen, Schaffer, and Windmeijer, 2020).

The underidentification test results are not specific to IV models, but apply to general settings of parameter matrices of potentially reduced rank, as discussed in Section 3. In this setting, Gospodinov, Kan, and Robotti (2017a) highlighted the problem that invariant rank tests for overidentification have no power to detect invalid moment conditions when the model is underidentified, and illustrated this problem in a Monte Carlo analysis within the setting of linear asset pricing factor models with the number of asset returns larger than the number of factors. We repeat this Monte Carlo analysis in Section 3.1, but now include results for tests for underidentification. We find that the underidentification tests perform well here, with the Sanderson-Windmeijer tests able to identify the spurious factors that cause underidentification. Section 3.1 further provides results for some of the applications in Gospodinov, Kan, and Robotti (2019).

The next section, Section 2 sets out the model and standard assumptions, and discusses the link between testing for over- and underidentification. Section 4 concludes. In Appendix B, it is shown that the relationships between rank tests, score tests and overidentification tests readily extend to testing for general rank, which establishes a direct link with the I-test for linear models as proposed by Arellano, Hansen, and Sentana (2012). Appendix C applies the methods to an asset pricing model example of Manresa, Peñaranda, and Sentana (2017). Further, Appendix D provides details of the limiting distributions of the Sargan tests for overidentification in underidentified homoskedastic models, as derived by Kitamura (2006) for the 2SLS estimator and Gospodinov et al. (2017a) for the LIML estimator. Appendix E contains proofs.

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 $P_A = A\left(A'A\right)^{-1}A'$ and $M_A = I_n - P_A$, with I_n the $n \times n$ identity matrix.

2 Tests for Underidentification

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 $k_z \geq k_x$. Note that exogenous explanatory variables, including the constant, 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 consider the following standard exclusion and relevance assumptions.

Assumption 1 $E(z_i u_i) = 0$.

Assumption 2 $E(z_i x_i') = Q_{zx}$; $\mathrm{r}(Q_{zx}) = k_x$.

Other maintained assumptions are such that IV estimators like 2SLS, LIML and GMM have standard limiting normal distributions when Assumptions 1 and 2 hold, and the data moments suffice to estimate the variances of these limiting distributions consistently. See e.g. Assumptions 3.2 and 3.5 in Hayashi (2000). These assumptions can hold for cross-sectional or time series data. As in Arellano et al. (2012) and Assumption 3.2 in Hayashi (2000), time series data are stationary and ergodic. We refer to the case where the observations are independent/not correlated and conditionally homoskedastic simply as the homoskedastic case. We defer discussion of a clustered design to Section 2.4 on dynamic panel data models.

Overidentification

Moment restriction $E(z_i u_i) = 0$, Assumption 1, is the exclusion restriction that, when $k_z > k_x$, is generally tested by a test of overidentifying restrictions, like the Hansen (1982) J-test. Tests for overidentifying restrictions are score tests for the hypothesis $H_0 : \gamma = 0$ in the IV model

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

using instruments Z, and where Z_o is any $k_z - k_x$ subset of instruments. The score test is invariant to the choice of instruments included in Z_o, unlike the Wald test, see Newey and West (1987) and the proof of Result 1 below in Appendix E. 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. Score and Wald tests have the same local asymptotic power for alternatives $\gamma = c/\sqrt{n}$, see e.g. Wooldridge (2010, p 417). It therefore follows that overidentification tests have the same local asymptotic power, independent of the estimator the overidentification test is based on.

Before discussing underidentification tests, we first show that the standard Hansen J-test based on the two-step GMM estimator is a robust score test for $H_0 : \gamma = 0$ in (3), based on a one-step GMM estimator for β and the OLS estimator for Π. This follows, but differs, from the result of Newey and West (1987), who considered the score test based on the efficient two-step GMM estimator. This observation will facilitate the exposition for the robust LIML based Kleibergen-Paap score/rank test for underidentification below.

A 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, wlog, $Z'H_{\widehat{u}_1}Z$ is an estimator for Ω_{zu} such that $n^{-1}Z'H_{\widehat{u}_1}Z \xrightarrow{p} \Omega_{zu}$ and Ω_{zu} is the variance of the limiting normal distribution of $\frac{1}{\sqrt{n}}Z'u \xrightarrow{d} N\left(0, \Omega_{zu}\right)$. For example, for a heteroskedasticity only robust specification, $H_{\widehat{u}_1} = \text{diag}\left(\widehat{u}_{1i}^2\right)$, an $n \times n$ diagonal matrix with the i-th diagonal element equal to $\widehat{u}_{1i}^2$.

Let $\widehat{u}_2 = y - X\widehat{\beta}_2$, then the Hansen J-statistic for overidentifying restrictions in model (1) is given by

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

Let $\widehat{\Pi} = \left(Z'Z\right)^{-1}Z'X$ be the OLS estimator of Π, and $\widehat{X} = Z\widehat{\Pi}$. Further let $\widehat{\gamma}$ be the 2SLS estimator of γ in the just identified model (3), given by

$$\widehat{\gamma} = \left(Z_o'M_{\widehat{X}}Z_o\right)^{-1} Z_o'M_{\widehat{X}}y$$

As Newey and West (1987) showed, $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right)$ is equal to the robust score test statistic based on the efficient two-step estimator, but also equal to a Wald type test statistic given by

$$W_{\gamma,\widehat{u}_1} = \widehat{\gamma}'\left(\widehat{Var}_{r,\widehat{u}_1}\left(\widehat{\gamma}\right)\right)^{-1}\widehat{\gamma}, \tag{5}$$

where the same estimator $Z'H_{\widehat{u}_1}Z/n$ of Ω_{zu} is used in the variance estimator. This results in a robust variance estimator based on the restricted, $\gamma = 0$, one-step GMM residual, and hence

$$W_{\gamma,\widehat{u}_1} = y'M_{\widehat{X}}Z_o\left(Z_o'M_{\widehat{X}}H_{\widehat{u}_1}M_{\widehat{X}}Z_o\right)^{-1}Z_o'M_{\widehat{X}}y.$$

It follows that $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right)$ is also equal to the robust score test statistic based on the one-step GMM estimator, which we state formally in the next result.

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

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

ii) Let the Hansen $J\left(\widehat{\beta}_2, \widehat{\beta}_1\right)$ be as defined in (4) and let $W_{\gamma,\widehat{u}_1}$ be as in (5). Then

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

Proof. ii) Follows from the results in Newey and West (1987) and the fact that $Z_o' M_{\widehat{X}} \widehat{u}_1 = Z_o' M_{\widehat{X}} y$. For i) and a general proof, see Appendix E. ■

It then follows that versions of the score test statistic for overidentifying restrictions can be obtained as Wald-type test statistics for $H_0 : \gamma = 0$ after OLS regression of the specification

$$\widehat{u}_1 = \widehat{X}\eta + Z_o\gamma + \xi, \tag{6}$$

see also Newey (1985), and using non-robust or robust estimators for $Var\left(\widehat{\gamma}\right)$ based either on $\widehat{u}_1$ or based on the residuals $\widehat{\xi} = M_Z \widehat{u}_1$. Under Assumptions 1 and 2 and the other maintained assumptions these test statistics have a limiting $\chi^2_{k_z - k_x}$ distribution, the non-robust versions valid in the homoskedastic case. The versions with the variance estimator based on $\widehat{u}_1$ are commonly called the "LM" versions, and those based on $\widehat{\xi}$ the "Wald" versions of the score test statistic, see e.g. the discussion in Baum, Schaffer, and Stillman (2007) and Bazzi and Clemens (2013). In the remainder, we will only consider the LM versions of the test statistics for clarity of exposition, but all have their Wald version counterparts.

Underidentification

Assumption 2, $\mathrm{r}\left(E\left(z_i x_i'\right)\right) = k_x$, is a necessary condition for the identification of β using the instrumental variables z_i. Let $E\left(z_i z_i'\right) = Q_{zz}$, with $\mathrm{r}\left(Q_{zz}\right) = k_z$. As $\Pi = Q_{zz}^{-1} E\left(z_i x_i'\right)$ it follows that the rank of Π is equal to the rank of $E\left(z_i x_i'\right)$. This means that β is identified iff $\mathrm{r}\left(\Pi\right) = k_x$. The model is therefore underidentified if $\mathrm{r}\left(\Pi\right) \leq k_x - 1$. Standard tests for underidentification, like the Cragg-Donald and Kleibergen-Paap tests are tests for $H_0 : \mathrm{r}\left(\Pi\right) = k_x - 1$ against $H_1 : \mathrm{r}\left(\Pi\right) = k_x$. One may restrict attention to the hypothesis $H_0 : \mathrm{r}\left(\Pi\right) = k_x - 1$, as the rejection probabilities of these rank test statistics are less than nominal size if $\mathrm{r}\left(\Pi\right) < k_x - 1$, see e.g. Gospodinov et al. (2017a) and the discussion in Section 3.1. For an alternative approach for testing $H_0 : \mathrm{r}\left(\Pi\right) \leq k_x - 1$, see Chen and Fang (2019).

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

$$\begin{aligned} x_1 &= Z\pi_1 + v_1 = Z\Pi_2\delta + v_1 \\ &= X_2\delta + v_1 - V_2\delta = X_2\delta + \varepsilon. \end{aligned} \tag{7}$$

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

$$E(z_i\varepsilon_i) = E(z_i(v_{1i} - v_{2i}'\delta)) = 0, \tag{8}$$

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

The intuition of orthogonality condition (8) is clear. If the instruments are not correlated with ε, then they have no explanatory power to predict x_1 after having controlled for the other endogenous explanatory variables in the model and the model is underidentified. Clearly, the moment condition $E(z_i\varepsilon_i) = 0$ in model (7) is an exclusion restriction like $E(z_i u_i) = 0$ in model (1), and likewise can be tested by standard tests for overidentification after estimation of the parameter vector δ by IV methods using instruments Z, which are score tests for $H_0 : \gamma = 0$ in the model specification

$$x_1 = X_2\delta + Z_o\gamma + \varepsilon, \tag{9}$$

where Z_o is any $k_z - k_x + 1$ subset of instruments, keeping the notation for γ and Z_o generic.

Below we show that the Cragg-Donald rank test statistic for testing $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$ for the homoskedastic case is identical to the Sargan (1958) test statistic for overidentifying restrictions in (7) after estimation of δ by LIML. The robust Cragg-Donald test statistic is equal to the Hansen J-statistic after estimation of δ by the Continuous Updating GMM estimator (CU-GMM). The standard version of the robust Kleibergen-Paap rank test statistic is equal to the robust score test statistic for testing $H_0 : \gamma = 0$ in (9) based on the LIML estimators for δ and Π_2, noting that the first-stage model for X_2 in (7) is given by $X_2 = Z\Pi_2 + V_2$.

The LIML and CU-GMM estimators are invariant to normalisation and therefore do not depend on the choice of x_1 as the dependent variable in (7), which is a necessary condition in order to equate the overidentification tests to rank tests. An overidentification test based on a non-invariant estimator, like two-step GMM, is therefore not a rank test.

However, if a rank test fails to reject the null, indicating that the model is underidentified, variable specific robust Hansen J-test statistics in the models $x_j = X_{-j}\delta_j + \varepsilon_j$, where X_{-j} is X without x_j, based on the GMM estimator of δ_j will provide additional information about which variable(s) are poorly predicted by the instruments. This is the robust extension of the variable specific conditional F-statistics of Sanderson and Windmeijer (2016).

Assumption 1, $E(z_i u_i) = 0$, clearly also implies a rank condition. Let $w_i = \begin{pmatrix} y_i & x_i' \end{pmatrix}'$, $\psi = \begin{pmatrix} 1 & -\beta' \end{pmatrix}'$ and $\Pi^* = Q_{zz}^{-1} E(z_i w_i)$. Then $E(z_i u_i) = E(z_i w_i')\psi = 0$ implies $\Pi^*\psi = 0$, and hence $\mathrm{r}(\Pi^*) = k_x$ when Assumption 2 holds and so $\mathrm{r}(\Pi) = k_x$. An invariant overidentification test is therefore a rank test for testing $H_0 : \mathrm{r}(\Pi^*) = k_x$ against $H_1 : \mathrm{r}(\Pi^*) = k_x + 1$. Note that $\Pi^* = \begin{bmatrix} \pi_y & \Pi \end{bmatrix}$ with π_y the reduced-form parameters

$$y = Z\Pi\beta + u - V\beta = Z\pi_y + v_y \tag{10}$$

It is therefore clear that the same tests for overidentifying restrictions that test Assumption 1 in the model structure $y = X\beta + u$, $X = Z\Pi + V$ are tests for Assumption 2 and hence underidentification when applied to the auxiliary model structure $x_1 = X_2\delta + \varepsilon$, $X_2 = Z\Pi_2 + V_2$. Whilst under Assumptions 1 and 2 the null for the overidentification test $E(z_i u_i) = 0$ holds, this is not the case for the null for the underidentification test.

2.1 Cragg-Donald Rank Test for Underidentification

The standard underidentification test is a test for $H_0 : \mathrm{r}(\Pi) = k_x - 1$ against $H_1 : \mathrm{r}(\Pi) = k_x$. Let $\pi = \mathrm{vec}(\Pi)$, $\pi_2 = \mathrm{vec}(\Pi_2)$ and $\widehat{\pi}$ and $\widehat{\pi}_2$ their OLS estimators. The Cragg and Donald (1993, 1997) rank test statistic, CD, is defined as

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

$$\text{s.t. } \mathrm{r}(\Pi) = k_x - 1,$$

or equivalently as the minimum of a nonlinear minimum distance criterion,

$$\begin{aligned} CD &= \min_{\delta,\Pi_2} MD(\delta, \Pi_2), \qquad (11) \\ MD(\delta, \Pi_2) &= \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \Pi_2\delta \\ \pi_2 \end{pmatrix}\right)' \left(V\widehat{a}r(\widehat{\pi})\right)^{-1} \left(\begin{pmatrix} \widehat{\pi}_1 \\ \widehat{\pi}_2 \end{pmatrix} - \begin{pmatrix} \Pi_2\delta \\ \pi_2 \end{pmatrix}\right). \end{aligned}$$

An estimator of $Var(\widehat{\pi})$ valid under homoskedasticity and the restriction that $\pi = 0$ is given by $V\widehat{a}r(\widehat{\pi}) = \widehat{\Sigma}_x \otimes (Z'Z)^{-1}$, where $\widehat{\Sigma}_x = X'X/n$. Let $V\widehat{a}r_r(\widehat{\pi})$ denote a robust variance estimator, also under the restriction that $\pi = 0$. Denote the resulting robust statistic by CD_r.

Let $\widehat{\delta}_L$ be the LIML estimator of δ, and $\widehat{\varepsilon}_L = x_1 - X_2\widehat{\delta}_L$. Then the Sargan (1958) test for overidentifying restrictions in model (7) is given by

$$\text{Sar}\left(\widehat{\delta}_L\right) = \frac{\widehat{\varepsilon}'_L P_Z \widehat{\varepsilon}_L}{\widehat{\varepsilon}'_L \widehat{\varepsilon}_L / n}.$$

Further, the CU-GMM estimator for δ in model (7) is defined as

$$\begin{aligned}
\widehat{\delta}_{cu} &= \arg\min_{\delta} J(\delta) \\
J(\delta) &= (x_1 - X_2\delta)' Z \left(Z' H_{\varepsilon(\delta)} Z\right)^{-1} Z' (x_1 - X_2\delta),
\end{aligned}$$

where $\varepsilon(\delta) = x_1 - X_2\delta$ and $Z'H_{\varepsilon(\delta)}Z$ is an efficient weight matrix taking into account serial correlation and/or heteroskedasticity. The Hansen J-statistic for overidentifying restrictions is then given by $J\left(\widehat{\delta}_{cu}\right)$.

Result 2 *Let CD be the Cragg-Donald statistic (11) for testing underidentification, $H_0 : \mathrm{r}(\Pi) = k_x - 1$, based on the homoskedastic variance estimator $\widehat{Var}(\widehat{\pi})$ and CD_r the test statistic based on the robust variance estimator $\widehat{Var}_r(\widehat{\pi})$. Then CD is equal to the Sargan test statistic for overidentifying restrictions in the auxiliary model (7), $CD = \text{Sar}\left(\widehat{\delta}_L\right)$, and CD_r is equal to the CU-GMM based J-statistic in the auxiliary model (7), $CD_r = J\left(\widehat{\delta}_{cu}\right)$.*

These results follow for LIML from the results in Hausman (1983) and Alonso-Borrego and Arellano (1999), and for CU-GMM from the derivations in Kleibergen and Mavroeidis (2009, Appendix), see also Gospodinov et al. (2017a). Arellano et al. (2012, p 261) already clarified the connection between the Cragg-Donald rank test statistic and the CU-GMM J-statistic, see Appendix B for further details.

2.2 Kleibergen-Paap Test

Let $\widehat{\Pi}_{2L}$ be the LIML estimator of Π_2 which can be obtained as

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

with, as before, $\widehat{\varepsilon}_L$ the LIML residual $\widehat{\varepsilon}_L = x_1 - X_2\widehat{\delta}_L$, see e.g. Godfrey and Wickens (1982). Let $\widehat{X}_{2L} = Z\widehat{\Pi}_{2L}$, then the robust version of the score test statistic for $H_0 : \gamma = 0$ in (9) evaluated at the LIML estimators for δ and Π_2 is given by

$$S_r\left(\widehat{\delta}_L\right) = \widehat{\varepsilon}'_L M_{\widehat{X}_{2L}} Z_o \left(Z'_o M_{\widehat{X}_{2L}} H_{\widehat{\varepsilon}_L} M_{\widehat{X}_{2L}} Z_o\right) Z'_o M_{\widehat{X}_{2L}} \widehat{\varepsilon}_L, \tag{12}$$

which can be seen as LIML based J-statistic equivalent from Result 1. By the invariance of LIML, this test is invariant to which explanatory variable is chosen as the dependent variable. $S_r\left(\widehat{\delta}_L\right)$ is the test statistic for testing $H_0 : \gamma = 0$ after OLS regression of the specification

$$\widehat{\varepsilon}_L = \widehat{X}_{2L}\eta + Z_o\gamma + \xi.$$

For the Kleibergen and Paap (2006) rank test, let G and F be $k_z \times k_z$ and $k_x \times k_x$ finite non-singular matrices respectively, and define $\Theta = G\Pi F'$ and $\widehat{\Theta} = G\widehat{\Pi}F'$. For testing $H_0 : \mathrm{r}(\Pi) = q$, 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 $k_x \times k_x$ 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 $\Theta = A_q B_q + A_{q,\perp}\Lambda_q B_{q,\perp}$, with $A_q' A_{q,\perp} = 0$ and $B_{q,\perp}B_q' = 0$. As $\Lambda_q = 0$ under the null $H_0 : \mathrm{r}(\Pi) = q$, the KP test is a test for $H_0 : \mathrm{vec}(\Lambda_q) = 0$.

The SVD applied to $\widehat{\Theta}$ yields the decomposition $\widehat{\Theta} = \widehat{A}_q\widehat{B}_q + \widehat{A}_{q,\perp}\widehat{\Lambda}_q\widehat{B}_{q,\perp}$; with $\widehat{\Lambda}_q = \widehat{A}_{q,\perp}'\widehat{\Theta}\widehat{B}_{q,\perp}'$, and the KP test statistic is given by

$$\mathrm{rk}(q) = \widehat{\lambda}_q'\widehat{\Omega}_q^{-1}\widehat{\lambda}_q, \tag{13}$$

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 $\widehat{\pi}$.

The next proposition gives the relationship between the Kleibergen-Paap test and the robust score test (12).

Proposition 1 *Consider the Kleibergen-Paap rank test statistic (13) for testing* $H_0 : \mathrm{r}(\Pi) = k_x - 1$. *Given choices of* F *and* G, *define the estimators* $\widehat{\beta}_{GF}$ *and* $\widehat{\Pi}_{2,GF}$ *as*

$$\left(\widehat{\delta}_{GF}, \widehat{\Pi}_{2,GF}\right) = \arg\min_{\delta,\Pi_2}\left(\begin{pmatrix}\widehat{\pi}_1\\ \widehat{\pi}_2\end{pmatrix} - \begin{pmatrix}\Pi_2\delta\\ \pi_2\end{pmatrix}\right)'(F'F \otimes G'G)\left(\begin{pmatrix}\widehat{\pi}_1\\ \widehat{\pi}_2\end{pmatrix} - \begin{pmatrix}\Pi_2\delta\\ \pi_2\end{pmatrix}\right).$$

Let $\widehat{\varepsilon}_{GF} = x_1 - X_2\widehat{\delta}_{GF}$, $\widehat{X}_{2,GF} = Z\widehat{\Pi}_{2,GF}$ *and let* Z_o *be a matrix of any* $k_z - k_x + 1$ *subset of instruments. Then*

$$\mathrm{rk}(k_x - 1) = \widehat{\varepsilon}_{GF}'M_{\widehat{X}_{2,GF}}Z_o\left(Z_o'M_{\widehat{X}_{2,GF}}H_{\widehat{\varepsilon}_{GF}}M_{\widehat{X}_{2,GF}}Z_o\right)^{-1}Z_o'M_{\widehat{X}_{2,GF}}\widehat{\varepsilon}_{GF}.$$

Proof. See Appendix E. ■

It follows from Proposition 1 that versions of $\text{rk}\,(k_x - 1)$ are test statistics for testing $H_0 : \gamma = 0$ in the specification

$$\widehat{\varepsilon}_{GF} = \widehat{X}_{2,GF}\eta + Z_o\gamma + \xi_{GF},$$

after estimation by OLS.

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

$$\widehat{\delta}_{GF} = \arg\min_{\delta} \frac{\psi' X' Z \left(Z'Z\right)^{-1} G'G \left(Z'Z\right)^{-1} Z'X\psi}{\psi' \left(F'F\right)^{-1} \psi}, \tag{14}$$

with $\psi = (1 \ \ -\delta')'$. It is clear that choosing F and G such that $F'F = \widehat{\Sigma}_x^{-1}$ and $G'G = \widehat{\Sigma}_z$ results in the LIML estimators for δ and Π_2. 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 test statistics commonly reported are based on the LIML normalisation. It therefore follows that the LIML based version of $\text{rk}\,(k_x - 1)$ is equal to $S_r\left(\widehat{\delta}_L\right)$ as defined in (12).

An invariant alternative to the KP test statistic is the J-statistic evaluated at an efficient two-step invariant LIML based estimator given by

$$\widehat{\delta}_{2L} = \left(\widehat{\Pi}_{2L}' Z'Z \left(Z'H_{\widehat{\varepsilon}_L} Z\right)^{-1} Z'X_2\right)^{-1} \widehat{\Pi}_{2L}' Z'Z \left(Z'H_{\widehat{\varepsilon}_L} Z\right)^{-1} Z'x, \tag{15}$$

with the invariant J-statistic calculated as,

$$J_{2L} \equiv J\left(\widehat{\delta}_{2L}\right) = \widehat{\varepsilon}_{2L}' Z \left(Z'H_{\widehat{\varepsilon}_{2L}} Z\right)^{-1} Z'\widehat{\varepsilon}_{2L}, \tag{16}$$

where $\widehat{\varepsilon}_{2L} = x_1 - X_2\widehat{\delta}_{2L}$, see Appendix A for details. An invariant efficient two-step estimator like $\widehat{\delta}_{2L}$ does not appear to have been considered before in the literature. Appendix A further considers an alternative two-step estimator based on the second step of an iterated GMM estimator that, if it converges, converges to the CU-GMM estimator.

2.3 Sanderson-Windmeijer Tests

In contrast to the invariant rank tests, 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 2SLS based versions of the underidentification tests. Let $\widehat{\delta}_j$ be the 2SLS estimator of δ_j in the model

$$x_j = X_{-j}\delta_j + \varepsilon_j \tag{17}$$

for $j=1,...,k_x$, using instruments Z, where X_{-j} is X without x_j. Let $\widehat{\varepsilon}_j = x_j - X_{-j}\widehat{\delta}_j$. The SW test statistics are given by

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

and SW provided the theory for testing for weak instruments based on the F-test versions, $F_j = \text{Sar}\left(\widehat{\delta}_j\right) / (k_z - k_x + 1)$. The weak instrument asymptotics they considered was that of $\text{r}(\Pi)$ local to a rank reduction of 1, or $\pi_j = \Pi_{-j}\delta_j + l/\sqrt{n}$. SW showed that the F_j statistics can provide additional information to that provided by the CD statistic about the nature of the weak instruments problem. The generalisation to robust tests for underidentification is then to compute the robust versions $S_r\left(\widehat{\delta}_j\right)$. From the results above it follows that these are simply the two-step Hansen J-statistics in (17), for $j = 1,...,k_x$.

2.4 Underidentification Tests 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. Following Holtz-Eakin et al. (1988), 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 x_{it} contains a lagged dependent variable and all other variables in x_{it} are endogenous, the available moment conditions at period t are given by

$$E\left(x_i^{t-2}\Delta u_{it}\right), \tag{18}$$

where $x_i^{t-2} = \begin{pmatrix} x_{i1}' & x_{i2}' & \dots & x_{i,t-2}' \end{pmatrix}'$. The moments (18) 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 $E\left(\text{vec}(V_i)\,\text{vec}(V_i)'\right) = \Sigma_{\text{vec}(V)}$. Whilst we can still make an assumption of conditional homoskedasticity, $E\left(\text{vec}(V_i)\,\text{vec}(V_i)' | Z_i\right) = \Sigma_{\text{vec}(V)}$, it seems implausible to assume the Kronecker structure $\Sigma_{\text{vec}(V)} = \Sigma_v \otimes I_n$. For example, due to the nature of the sequential moments, the variances $E\left(v_{ijt}^2\right)$ are likely to be varying over time, as are the covariances $E\left(v_{ijt} v_{ist'}\right)$.

Therefore, the non-robust version of the CD test statistic for testing $H_0 : \text{r}(\Pi) = k_x - 1$ 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 is no longer a simple minimum eigenvalue problem and the solution needs to be obtained via iterative methods.

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

$$\Delta x_{1i} = (\Delta X_{2i})\,\delta + \varepsilon_{1i}, \tag{19}$$

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

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

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

The LIML normalised cluster-robust Kleibergen-Paap test is based on the pooled LIML estimators of δ in (19) and Π_2 in $\Delta X_{2i} = Z_i \Pi_2 + V_{2i}$. Under the null that $\text{r}(\Pi) = k_x - 1$, these test statistics are invariant and have a limiting $\chi^2_{k_z - k_x + 1}$ distribution under the maintained assumptions.

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 $H_0 : E\left(Z_i'\varepsilon_{j,i}\right) = 0$ with the Hansen J-statistic. These tests are therefore easily implemented using standard estimation routines for the Arellano-Bond GMM estimator.

The above methods extend straightforwardly to the System estimator of Blundell and Bond (1998), and, instead of first differences, one could specify the transformation as that of orthogonal deviations, see Arellano and Bover (1995).

Arellano, Hansen, and Sentana (1999, 2012) also considered testing for underidentification in dynamic panel data models using their I-test. A general discussion of the I-test is presented in Appendix B. They considered in particular an AR(2) specification in first-differences,

$$\Delta y_{it} = \alpha_1 \Delta y_{i,t-1} + \alpha_2 \Delta y_{i,t-2} + \Delta u_{it}.$$

A comparison of the approach to testing for underidentification followed here and the I-test for this model is presented in Appendix B.2.

2.4.1 Some Monte Carlo Results

Before we apply the underidentification tests to an application from Acemoglu et al. (2008) below, we present some Monte Carlo results for a related design. The data are generated from

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{i,t-1} + \eta_i + u_{it} \\ x_{it} &= \rho x_{i,t-1} + (1-\rho)\,\eta_i + \theta u_{it} + w_{it}, \end{aligned}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, with η_i, u_{it} and w_{it} independent $N(0,1)$ distributed random variables, $y_{i0} \sim N(0,1)$ and $x_{i0} \sim N(0,1)$, $\theta = 0.25$ and $\alpha = \beta = 0.5$, with therefore the long-run effect equal to $\beta/(1-\alpha) = 1$. We further set $T = 6$, and consider values of $n = 100$, 500 and 1000. Sequential instruments used at time t are $y_{i1}, x_{i,1}, \ldots, y_{i,t-2}, x_{i,t-2}$, resulting in a total of $k_z = 20$ instruments. When $\rho = 1$, we have that $\Delta x_{i,t-1} = \theta u_{i,t-1} + w_{i,t-1}$ and the model is underidentified as the instruments have then no predictive power for $\Delta x_{i,t-1}$. We present two-step Arellano-Bond estimation results and rejection frequencies of the underidentification tests at the 5% level from $10,000$ Monte Carlo replications in Table 1, for $\rho = 1$, but also for $\rho = 0.9$ and $\rho = 0.7$.

For the underidentified specification, the rejection frequencies of the test statistics are slightly below the 5% level for the small sample size $n = 100$, more so for the robust CD test statistic at 0.025 than the for the robust KP and J_{2L} statistics at 0.036 and

0.035 respectively. The SW test statistics correctly convey the information that the instruments fail to predict $\Delta x_{i,t-1}$, but do have predictive power for $\Delta y_{i,t-1}$. For $n = 500$ and $n = 1000$, the test statistics have correct size.

Table 1: Monte Carlo results

	α	β	$\frac{\beta}{1-\alpha}$	KP	J_{2L}	CD	SW	
$n = 100$							$\Delta y_{i,t-1}$	$\Delta x_{i,t-1}$
$\rho = 1$	0.440 (0.103)	0.356 (0.305)	0.650 (0.568)	0.036	0.035	0.025	0.883	0.038
$\rho = 0.9$	0.380 (0.121)	0.332 (0.235)	0.598 (0.477)	0.095	0.095	0.084	0.618	0.100
$\rho = 0.7$	0.367 (0.133)	0.397 (0.152)	0.695 (0.393)	0.298	0.302	0.288	0.688	0.380
$n = 500$								
$\rho = 1$	0.488 (0.044)	0.335 (0.287)	0.657 (0.564)	0.046	0.046	0.046	0.953	0.046
$\rho = 0.9$	0.456 (0.066)	0.417 (0.142)	0.802 (0.349)	0.767	0.767	0.766	0.983	0.674
$\rho = 0.7$	0.463 (0.067)	0.469 (0.069)	0.899 (0.235)	0.999	0.999	0.999	0.999	0.999
$n = 1000$								
$\rho = 1$	0.494 (0.031)	0.336 (0.284)	0.666 (0.564)	0.052	0.052	0.052	0.960	0.052
$\rho = 0.9$	0.476 (0.049)	0.454 (0.103)	0.888 (0.271)	0.991	0.991	0.991	0.999	0.982
$\rho = 0.7$	0.481 (0.048)	0.485 (0.049)	0.948 (0.173)	1.000	1.000	1.000	1.000	1.000

Notes: Mean and (sd) of two-step Arellano-Bond GMM estimates, and rejection frequencies of robust underidentification tests at 5% level from 10,000 replications. $T = 6$, $\alpha = 0.5$, $\beta = 0.5$. $KP = S_r\left(\widehat{\delta}_L\right)$, $J_{2L} = J\left(\widehat{\delta}_{2L}\right)$, $CD = J\left(\widehat{\delta}_{cu}\right)$, $SW_j = J\left(\widehat{\delta}_j\right)$, shorthand for $J\left(\widehat{\delta}_{j2}, \widehat{\delta}_{j1}\right)$.

For $\rho < 1$, the Arellano-Bond estimator is consistent, but as the results in the table show, its finite-sample behaviour is poor for $\rho = 0.9$ and $n = 100$. This is reflected in a low rejection probability of around 10% for the underidentification test statistics. The estimation results are still quite poor for $\rho = 0.7$ and $n = 100$, with the bias of the long-run effect not much different from the unidentified case. The rejection frequencies of the underidentification test statistics is here around 30%. For the larger sample sizes of $n = 500$ and $n = 1000$, the different underidentification test statistics have identical rejection frequencies. The estimation results improve with increasing n, with the best results for $n = 1000$ and $\rho = 0.7$, as expected. The underidentification tests in this latter case have a rejection frequency of 100% and the average p-values of the test statistics are then equal to 0.000. As is the case with the standard F-statistic for testing for underidentification, a rejection of the null of underidentification does not necessarily imply that the estimator is well behaved and has a small bias. However, it is clear that when the null of underidentification is not rejected, estimation results are not reliable.

2.4.2 An Application

We investigate underidentification in an example taken from Acemoglu, Johnson, Robinson, and Yared (2008) (AJRY), as recently revisited by Hansen and Lee (2020). AJRY consider the effect of income on democracy and specify the linear dynamic panel data model

$$d_{it} = \alpha d_{i,t-1} + \beta inc_{i,t-1} + \tau_t + \eta_i + u_{it},$$

where d_{it} is a measure of democracy and inc_{it} is log income per capita. The data used is a panel of 127 countries observed over the period 1960-2000 at 5-year and 10-year frequencies. Following Hansen and Lee (2020), we consider the AJRY estimates as reported in their Table 2. The parameters are estimated using the Arellano-Bond GMM first-differenced estimator, with the instruments specified as

$$Z_i = \begin{bmatrix} d_{i1} & 0 & 0 & 0 & inc_{i1} \\ 0 & d_{i1}\, d_{i2} & 0 & 0 & inc_{i2} \\ 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & d_{i1} \dots d_{i,T-2} & inc_{i,T-2} \end{bmatrix}.$$

Hansen and Lee (2020) found large differences between the one-step Arellano-Bond GMM estimator and the iterated GMM estimator, which iterates over the weight matrix until convergence, for the 5-year frequency data, but not for the 10-year frequency data. These results are presented in Table 2.

Table 2: Estimation results, Arellano-Bond estimator

	one-step GMM			iterated GMM		
Frequency	α	β	$\frac{\beta}{1-\alpha}$	α	β	$\frac{\beta}{1-\alpha}$
5 year	0.489	-0.129	-0.253	0.744	-0.009	-0.036
$k_z - k_x = 44$	(0.095)	(0.088)	(0.163)	(0.128)	(0.039)	(0.149)
$n = 127$, $\#obs = 838$						
10 year	0.227	-0.318	-0.411	0.288	-0.280	-0.393
$k_z - k_x = 15$	(0.125)	(0.183)	(0.246)	(0.146)	(0.202)	(0.290)
$n = 118$, $\#obs = 338$						

Notes: From Acemoglu et al. (2008, Table 2) and Hansen and Lee (2020, Table 4). Misspecification-robust standard errors of Hansen and Lee (2020) in brackets

This sensitivity to the choice of initial weight matrix could be an indication of underidentification. The results in Table 3 confirm this. The robust KP, J_{2L} and CD test statistics do not reject the null of underidentification for this specification using the 5-year frequency data, with p-values of 0.135, 0.132 and 0.231 respectively.[2] The SW statistics show that especially the lag of log income per capita in the differenced model, $\Delta inc_{i,t-1}$ is poorly predicted by the instruments. The p-values of the robust test statistics for the 10-year frequency data are smaller, at 0.023, 0.024 and 0.026, but therefore do not reject the null of underidentification at the 1% level, with the SW statistic for $\Delta inc_{i,t-1}$ having a p-value of 0.022, whereas the SW p-value for $\Delta d_{i,t-1}$ is equal to 0.000.

Table 3: P-values of robust underidentification test statistics

	KP	J_{2L}	CD	SW	
Frequency				$\Delta d_{i,t-1}$	$\Delta inc_{i,t-1}$
5 year	0.135	0.132	0.231	0.018	0.121
10 year	0.023	0.024	0.026	0.000	0.022

3 Testing the Rank of General Parameter Matrices

The LIML and CU-GMM based tests of rank of the matrices Π^* and Π may appear to be specific to the linear IV setup, with Z used as instruments for both the over- and underidentification tests. This approach can, however, be applied to more general settings testing the rank of parameter matrices. Consider the possibly reduced rank linear model

[2]These results can be obtained using the UNDERID routine (Schaffer and Windmeijer, 2020) in Stata.

specification as considered in Anderson (1951), see also Reinsel and Velu (1998) and 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 $Y = \begin{bmatrix} y_1 & Y_2 \end{bmatrix}$, $B = \begin{bmatrix} b_1 & B_2 \end{bmatrix}$, $\widehat{B} = \begin{bmatrix} \widehat{b}_1 & \widehat{B}_2 \end{bmatrix}$ and $X = \begin{bmatrix} X_1 & X_o \end{bmatrix}$, where X_o 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

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

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

$$y_1 = Y_2 \delta + \varepsilon_1, \tag{20}$$

using X as instruments. 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 test statistics can again be obtained as the Wald-type test statistics for $H_0 : \gamma = 0$ in the specification

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

estimated by OLS. The robust CD test statistic is $J\left(\widehat{\delta}_{cu}\right)$ after estimation of (20) by CU-GMM. The extensions to general rank tests, $H_0 : \mathrm{r}(B) = q$, are then as discussed in Appendix B.

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 $Var\left(\mathrm{vec}\left(\widehat{B}'\right)\right)$ 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 test statistics for, for example, $H_0 : \mathrm{r}(B') = k_x - 1$ are identical to the rank test statistics for $H_0 : \mathrm{r}(C) = k_x - 1$. Partition $X = \begin{bmatrix} x_1 & X_2 \end{bmatrix}$, then the test statistics 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 asset pricing models as discussed next.

3.1 Asset-Pricing Models

Gospodinov, Kan, and Robotti (2017a) (GKR) considered the behaviour of the CU-GMM J-test for overidentification 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

$$E\left(R_t x_t'\lambda - \iota_{k_R}\right) = 0. \tag{21}$$

Let $e_t(\lambda) = R_t x_t'\lambda - \iota_{k_R}$. 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 $H = \begin{bmatrix} \iota_{k_R} & E(R_t x_t') \end{bmatrix}$, then (21) can be written as $H\psi = 0$, with $\psi = \begin{pmatrix} -1 & \lambda' \end{pmatrix}'$, and GKR showed that therefore testing moment conditions (21) is equivalent to testing $H_0 : \text{r}(H) = k_x$ against $H_1 : \text{r}(H) = k_x + 1$. From this it follows that if $\text{r}(E(R_t x_t')) < k_x$, the CU-GMM J-test has no power to detect violations of the moment conditions (21).

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 $P_1 P_1' = I_{k_R} - \iota_{k_R}\iota_{k_R}'/k_R$. Let $R_{1t} = P_1' R_t$, then GKR show that the $J\left(\widehat{\lambda}_{cu}\right)$-statistic is the same as the robust Cragg-Donald statistic for testing $H_0 : \text{r}(E(R_{1t}x_t')) = k_x - 1$ against $H_1 : \text{r}(E(R_{1t}x_t')) = k_x$.

The latter formulation fits the testing procedure described in Section 3. The CD test for $H_0 : \text{r}(E(R_{1t}x_t')) = k_x - 1$ is the test for $H_0 : \text{r}(\Lambda_1') = k_x - 1$ 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 hypothesis can then be reformulated as $H_0 : \text{r}(C_1) = k_x - 1$ in the specification

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

With $X = \begin{bmatrix} \iota_T & F \end{bmatrix}$, we can therefore obtain robust invariant test statistics for overidentification by estimating the specification

$$\iota_T = F\delta + \varepsilon \tag{22}$$

by LIML or CU-GMM, using the $T \times (k_R - 1)$ matrix $R_1 = [R_{1t}]$ as instruments for F.

GKR confirmed in a Monte Carlo study the poor performance of the overidentification test in underidentified models. Here, the model is underidentified if $\text{r}(E(R_t x_t')) < k_x$ 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 $E(R_t x_t')$ itself. These tests are the same as the overidentification tests above for model (22), but now with the $T \times k_R$ matrix $R = [R_t]$ used as instruments instead of R_1.

Tables 4 and 5 present some results of 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 $T = 200$ and $T = 1000$. 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, and include the non-robust Sargan statistic as well for comparison. Unlike GKR, we do not take deviations from the means of the residuals when computing the variance-covariance matrix of the moments.

Table 4: Rejection frequencies of overidentification test statistics at 5% level

T	# spur factors	Sar $\left(\widehat{\lambda}_L\right)$	KP, S_r $\left(\widehat{\lambda}_L\right)$	J_{2L}, J $\left(\widehat{\lambda}_{2L}\right)$	CD, J $\left(\widehat{\lambda}_{cu}\right)$	J $\left(\widehat{\lambda}_{hj}\right)$
200	0	0.4848	0.4309	0.4234	0.3448	0.5604
	1	0.0117	0.0112	0.0092	0.0019	0.4793
	2	0.0002	0.0005	0.0004	0.0000	0.3909
1000	0	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0424	0.0382	0.0377	0.0371	0.9808
	2	0.0012	0.0008	0.0008	0.0007	0.9309

Notes: Design as in GKR Table 1, moment conditions (21) are invalid. 10,000 MC replications

Table 5: Rejection frequencies of underidentification test statistics at 5% level

T	# spur factors	Sar $\left(\widehat{\delta}_L\right)$	KP, S_r $\left(\widehat{\delta}_L\right)$	J_{2L}, J $\left(\widehat{\delta}_{2L}\right)$	CD, J $\left(\widehat{\delta}_{cu}\right)$
200	0	0.9991	0.9840	0.9826	0.9778
	1	0.0396	0.0275	0.0241	0.0138
	2	0.0013	0.0015	0.0011	0.0002
1000	0	1.0000	1.0000	1.0000	1.0000
	1	0.0487	0.0433	0.0429	0.0428
	2	0.0016	0.0018	0.0017	0.0012

As in GKR, the results in Table 4 confirm the poor power properties of the invariant test statistics for overidentifying restrictions in underidentified models. Table 4 also presents the results for a two-step Hansen J-statistic, denoted for brevity $J\left(\widehat{\lambda}_{hj}\right)$, with the one-step estimator the one that minimises the Hansen and Jagannathan (1997) distance,

$$\widehat{\lambda}_{hj} = T\left(X'R\left(R'R\right)^{-1}R'X\right)^{-1}X'R\left(R'R\right)^{-1}\iota_{k_R},$$

confirming that this test does retain power to reject the false null in underidentified models like the standard 2SLS-based Sargan test, as shown in Appendix D.

For the homoskedastic case, Guggenberger, Kleibergen, and Mavroeidis (2019) propose a conditional subvector Anderson-Rubin test procedure that results in more power than when using standard asymptotic chi-squared critical values, whilst controlling size. This test can be seen as a LIML based overidentification test, and the conditional approach of Guggenberger et al. (2019) can be applied to the Sargan test for overidentification, $\text{Sar}\left(\widehat{\lambda}_L\right)$. However, in underidentified models as considered here, this conditional test procedure will not have power exceeding size, see Guggenberger et al. (2019, p 497), and so will not lead to meaningfully improved power in these underidentified models.

The limiting distribution results for the invariant overidentification test statistics as derived by GKR also apply to the invariant underidentification test statistics, meaning that the test statistics converge in distribution to a $\chi_{k_r-k_x+1}$ under the null that $\text{r}\left(E\left(R_t x_t'\right)\right) = k_x - 1$ and the maintained assumptions. If $\text{r}\left(E\left(R_t x_t'\right)\right) < k_x - 1$ then the rejection frequency will be less than size. This is confirmed by the results in Table 5. In this design, the underidentification tests correctly convey that the model is underidentified.

Table 6 presents the results for the robust SW $J\left(\widehat{\delta}_j\right)$ statistics, for the models $x_j = X_{-j}\delta_j + \varepsilon_j$, the one-step estimator being 2SLS. For this design, these test statistics give a clear indication of which factors are the spurious ones, with low rejection frequencies for the spurious factors f_4 and f_5.

Table 6: Rejection frequencies of robust SW $J\left(\widehat{\delta}_j\right)$ statistics at 5% level

T	# spur factors	ι	f_1	f_2	f_3	f_4	f_5
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

Gospodinov, Kan, and Robotti (2019) highlighted the problem of the presence of spurious factors in the asset pricing models as proposed by Jagannathan and Wang (1996) (C-LAB), that includes as risk factors the market excess returns (vw), the growth rate in per capita labor income ($labor$) and the lagged default premium ($prem$), which is the yield spread between Baa- and Aaa-rated corporate bonds, and as proposed by Lettau and Ludvigson (2001) (CC-CAY), that includes the growth rate in real per capita nondurable consumption (cg), the lagged consumption-aggregate wealth ratio (cay) and an interaction term between cg and cay ($cg \cdot cay$) as risk factors.

Table 7 presents the p-values for the uncentered robust over- and underidentification

Table 7: P-values of robust over- and underidentification test statistics

	overidentification				underidentification		
	KP	J_{2L}	CD	$J\left(\widehat{\lambda}_{hj}\right)$	KP	J_{2L}	CD
C-LAB	0.389	0.406	0.684	0.000	0.438	0.447	0.700
CC-CAY	0.936	0.921	0.949	0.001	0.944	0.947	0.956

Notes: Models and data from Gospodinov et al. (2019), Tables 5 and 6, Panels A.

Table 8: P-values of SW $J\left(\widehat{\delta}_j\right)$ statistics

	ι	vw	$labor$	$prem$	cg	cay	$cg \cdot cay$
C-LAB	0.420	0.000	0.394	0.661			
CC-CAY	0.000				0.567	0.487	0.938

test-statistics for the two models, using the data as in Gospodinov et al. (2019) that consist of monthly value-weighted gross returns on 25 Fama-French size and book-to-market ranked portfolios for the $T = 552$ months January 1967 to December 2012. Centered CD test statistics were reported in Gospodinov et al. (2017b, 2019). The KP and J_{2L} statistics further confirm the problem of underidentification in these two asset pricing models. As expected, the invariant overidentification tests statistics do not indicate any misspecification problems. However, the two-step overidentification test statistic $J\left(\widehat{\lambda}_{hj}\right)$ indicates that both models are misspecified in the sense that moment conditions (21) are not valid, a conclusion also reached by Gospodinov et al. (2017b) on the bases of the Hansen and Jagannathan (1997) distance statistic.

The SW variable specific underidentification statistics provide interesting additional information, as displayed in Table 8. For the C-LAB model, only vw is correlated with the returns R, whereas $labor$ and $prem$ could be spurious factors. For the CC-CAY model all factors cg, cay and $cg \cdot cay$ could be spurious. These results line up exactly with the model selection in Gospodinov et al. (2019) that selects vw as the only remaining factor for the C-LAB model and no factors remaining for the CC-CAY model, with for both these models the overidentification test statistics rejecting the moment conditions (21).

Following arguments similar to those in Manresa, Peñaranda, and Sentana (2017), the problems with these specifications can directly be inferred from the ranks of the matrices $H = \left[\begin{array}{cc} \iota_{k_R} & E\left(R_t x_t'\right) \end{array}\right]$ and $E\left(R_t x_t'\right)$.[3] For a combination of the factors to form

[3]Appendix C performs related rank tests for one of the specifications for excess returns in Manresa et al. (2017).

Table 9: P-values of robust rank test statistics for H and $E\left(R_t x_t'\right)$

		H			$E\left(R_t x_t'\right)$		
	rank reduction	KP	J_{2L}	CD	KP	J_{2L}	CD
C-LAB	2	0.592	0.591	0.602	0.519	0.519	0.521
	3	0.004	0.005	0.007	0.000	0.000	0.000
CC-CAY	2	0.965	0.962	0.968	0.949	0.946	0.956
	3	0.922	0.920	0.924	0.934	0.934	0.934
	4	0.026	0.026	0.026	0.000	0.000	0.000

a meaningful pricing SDF, it should be the case that $\mathrm{r}(H) = \mathrm{r}(E(R_t x_t')) > 0$. Table 9 provides results for sequential rank tests. A rank reduction of $j > 0$ indicates testing $H_0 : \mathrm{r}(H) = 1 + k_x - j$ against $H_1 : \mathrm{r}(H) = 2 + k_x - j$, and $H_0 : \mathrm{r}(E(R_t x_t')) = k_x - j$ against $H_1 : \mathrm{r}(E(R_t x_t')) = 1 + k_x - j$. It is clear that for both the C-LAB and CC-CAY models, the reduction in rank in H is driven by the reduction in rank of $E(R_t x_t')$. The results indicate that $\mathrm{r}(H) = 3$, $\mathrm{r}(E(R_t x_t')) = 2$ for the C-LAB model and $\mathrm{r}(H) = 2$, $\mathrm{r}(E(R_t x_t')) = 1$ for the CC-CAY model, and hence neither provide meaningful SDFs, which is the same conclusion reached above and by Gospodinov et al. (2019).

4 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), although Windmeijer (2019) provides a link between robust underidentification tests and the behaviour of a GMM estimator where the weight matrix is based on the first-stage residuals. Whereas a rejection of the null of underidentification does not necessarily imply strong instruments, 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.

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 based tests are relatively insensitive to weak identification, see Newey and Windmeijer (2009). For cross-sectional models with heteroskedasticity, the proposal of Hausman, Newey, Woutersen, Chao, and Swanson (2012) and Chao, Hausman, Newey, Swanson, and Woutersen (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 auxiliary linear model.

Acknowledgements

I would like to thank two referees and an associate editor, Peter Boswijk, Jason Dovonon, Jean-Marie Dufour, Bo Honoré, Pascal Lavergne, Whitney Newey, Chris Muris, Mark Schaffer, Enrique Sentana, Jon Temple and seminar and conference participants at Amsterdam, Bristol, Canterbury, Durham, Hong Kong, Lisbon, Liverpool, Melbourne, Montreal, Oxford, Perth and Toulouse for helpful comments.

Supplementary Materials, Appendices A-E

References

Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2008): "Income and Democracy," *American Economic Review*, 98, 808–842.

Al-Sadoon, M. M. (2017): "A Unifying Theory of Tests of Rank," *Journal of Econometrics*, 199, 49–62.

Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 17, 36–49.

Anderson, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *The Annals of Mathematical Statistics*, 22, 327–351.

Andrews, I. (2018): "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, 100, 337–348.

Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 58, 277.

Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

Arellano, M., L. P. Hansen, and E. Sentana (1999): "Underidentification?" Mimeo, CEMFI, ftp://ftp.cemfi.es/pdf/papers/es/Ahsc.pdf.

——— (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, 650–659.

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.

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.

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.

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.

Chen, Q. and Z. Fang (2019): "Improved Inference on the Rank of a Matrix," *Quantitative Economics*, 10, 1787–1824.

Cragg, J. G. and S. G. Donald (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

——— (1997): "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250.

Godfrey, L. and M. Wickens (1982): "A Simple Derivation of the Limited Information Maximum Likelihood Estimator," *Economics Letters*, 10, 277–283.

Gospodinov, N., R. Kan, and C. Robotti (2017a): "Spurious Inference in Reduced-Rank Asset-Pricing Models," *Econometrica*, 85, 1613–1628.

——— (2017b): "Too Good to Be True? Fallacies in Evaluating Risk Factor Models," Working Paper Series 2017-9, Federal Reserve Bank of Atlanta.

——— (2019): "Too Good to Be True? Fallacies in Evaluating Risk Factor Models," *Journal of Financial Economics*, 132, 451–471.

Guggenberger, P., F. Kleibergen, and S. Mavroeidis (2019): "A More Powerful Subvector Anderson Rubin Test in Linear Instrumental Variables Regression," *Quantitative Economics*, 10, 487–526.

Hansen, B. E. and S. Lee (2020): "Inference for Iterated GMM Under Misspecification," Mimeo, University of Wisconsin, ssc.wisc.edu/ bhansen/papers/IteratedGMM.pdf.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029.

Hansen, L. P. and R. Jagannathan (1997): "Assessing Specification Errors in Stochastic Discount Factor Models," *The Journal of Finance*, 52, 557–590.

Hausman, J. A. (1983): "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, Elsevier, vol. I, 391–448.

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 (2019): "Further Results on the Weak Instruments Problem of the System GMM Estimator in Dynamic Panel Data Models," *Oxford Bulletin of Economics and Statistics*, 82, 453–481.

Hayashi, F. (2000): *Econometrics*, Princeton University Press.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56, 1371.

Jagannathan, R. and Z. Wang (1996): "The Conditional CAPM and the Cross-Section of Expected Returns," *The Journal of Finance*, 51, 3–53.

Kim, U. (2019): "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. (2006): "Specification Tests with Instrumental Variables and Rank Deficiency," in *Econometric Theory and Practice*, ed. by D. Corbae, S. N. Durlauf, and B. E. Hansen, Cambridge University Press, 59–81.

KLEIBERGEN, F. AND S. MAVROEIDIS (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27, 293–311.

KLEIBERGEN, F. AND R. PAAP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

KLEIBERGEN, F., M. E. SCHAFFER, AND F. WINDMEIJER (2020): "RANKTEST: Stata Module to Test the Rank of a Matrix," https://ideas.repec.org/c/boc/bocode/s456865.html.

LETTAU, M. AND S. LUDVIGSON (2001): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238–1287.

MANRESA, E., F. PEÑARANDA, AND E. SENTANA (2017): "Empirical Evaluation of Overspecified Asset Pricing Models," CEPR Discussion Paper Series DP12085, CEPR, London.

NEWEY, W. K. (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

NEWEY, W. K. AND K. D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777.

NEWEY, W. K. AND F. WINDMEIJER (2009): "Generalized Method of Moments with Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

REINSEL, G. C. AND R. P. VELU (1998): *Multivariate Reduced-Rank Regression*, Springer New York.

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.

SCHAFFER, M. E. AND F. WINDMEIJER (2020): "UNDERID: Postestimation Tests of Under- and Over-Identification After Linear IV Estimation," https://ideas.repec.org/c/boc/bocode/s458805.html.

STOCK, J. H. AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models*, ed. by D. W. K.

Andrews and J. H. Stock, Cambridge University Press, 80–108.

Windmeijer, F. (2019): "Weak Instruments, First-Stage Heteroskedasticity and the Robust F-Test," Dept. of Economics Discussion Paper 19/708, University of Bristol.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge and London, second ed.

Supplementary Appendices for "Testing Underidentification in Linear Models, with Applications to Dynamic Panel and Asset Pricing Models"

Frank Windmeijer
Department of Statistics and Nuffield College
University of Oxford, UK
and Cemmap, UCL/IFS, London, UK

March 19, 2021

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

For estimating β in structural model (1) with first-stage specification (2), if Π were known, then the natural just-identifying linear combination of instruments 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, where $Q_{zz} = \text{plim}\,(Z'Z/n)$,

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

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

An optimal invariant two-step estimator based on the LIML estimators $\widehat{\beta}_L$ and $\widehat{\Pi}_L$ is given by

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

$\widehat{\beta}_{2L}$ has the same limiting distribution as the optimal infeasible estimator (A.1) and is invariant to normalisation. The Hansen J-statistic calculated as

$$J\left(\widehat{\beta}_{2L}\right) = \widehat{u}_{2L}' Z \left(Z'H_{\widehat{u}_{2L}}Z\right)^{-1} Z'\widehat{u}_{2L}, \quad \text{(A.3)}$$

with $\widehat{u}_{2L} = y - X\widehat{\beta}_{2L}$ is also invariant to normalisation, and $J\left(\widehat{\beta}_{2L}\right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under Assumptions 1 and 2 and further maintained assumptions.

A.1 CU-GMM as an Iterated GMM Estimator

Let $\Pi^* = \left[\begin{array}{cc} \pi_y & \Pi \end{array}\right]$, $\widehat{\Pi}^* = \left[\begin{array}{cc} \widehat{\pi}_y & \widehat{\Pi} \end{array}\right]$, $\pi = \text{vec}\left(\Pi\right)$, $\widehat{\pi} = \text{vec}\left(\widehat{\Pi}\right)$, and $\widehat{\pi}^* = \text{vec}\left(\widehat{\Pi}^*\right)$. For the CU-GMM estimator we have the following result.

Lemma A.1 *Consider the CU-GMM minimum distance estimators*

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

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 E. ■

From Lemma A.1 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 $\widehat{\Pi}$.

From the first-order condition for the CU-GMM estimator of Π, we have for $\widehat{\Pi}_{cu}$,

$$\text{vec}\left(\widehat{\Pi}_{cu}\right) = \left(\widehat{C}_{cu}' \left(V\widehat{a}r_r\left(\widehat{\pi}^*\right)\right)^{-1} \widehat{C}_{cu}\right)^{-1} \widehat{C}_{cu}' \left(V\widehat{a}r_r\left(\widehat{\pi}^*\right)\right)^{-1} \widehat{\pi}^*, \quad \text{(A.4)}$$

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 and $\widehat{u}_1 = y - X\widehat{\beta}_1$. Let $\widehat{C}_1 = \left[\begin{array}{cc} \widehat{\beta}_1 & I_{k_x} \end{array}\right]' \otimes I_{k_z}$, and

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

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 $\widehat{\Pi}_{cu}$ if the scheme converges.

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 (A.2) 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

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

with $\widehat{C}_L = \begin{bmatrix} \widehat{\beta}_L & I_{k_x} \end{bmatrix}' \otimes I_{k_z}$, and 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 $\widehat{u}_{2L,r} = y - X\widehat{\beta}_{2L,r}$.

B Testing for General Rank and the Arellano-Hansen-Sentana I Test

In order to provide the link to the Arellano, Hansen, and Sentana (2012) (AHS) test, we return to the original structural and first-stage models (1) and (2). Specify $w_i = \begin{pmatrix} y_i & x_i' \end{pmatrix}'$ 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

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

or $r\left(\Pi^*\right) = k_x - 1$. As in AHS, we start by standardising $\Psi = \begin{bmatrix} I_2 \\ \Delta \end{bmatrix}$. Partition $w = \begin{bmatrix} y & x_1 & X_2 \end{bmatrix}$ and $\Pi^* = \begin{bmatrix} \pi_y & \pi_1 & \Pi_2 \end{bmatrix}$. We then have the two equations

$$y = X_2\delta_y + \varepsilon_y \quad \text{(B.1)}$$

$$x_1 = X_2\delta_x + \varepsilon_x \quad \text{(B.2)}$$

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

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

$$\begin{aligned} y &= X_2\delta_y + Z_o\gamma_y + \varepsilon_y \\ x_1 &= X_2\delta_x + Z_o\gamma_x + \varepsilon_x, \end{aligned}$$

then a test for $H_0 : E(z_i\varepsilon_i') = 0$ 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 $\begin{pmatrix} \widehat{\gamma}_y & \widehat{\gamma}_x \end{pmatrix} = \left(Z_o'M_{\widehat{X}_2}Z_o\right)^{-1} Z_o'M_{\widehat{X}_2} \begin{pmatrix} y & x_1 \end{pmatrix}$, where $\widehat{X}_2 = Z\widehat{\Pi}_2$. These IV estimators are also efficient under conditional homoskedasticity, $Var(\varepsilon_i|z_i) = \Sigma_\varepsilon$, by standard SURE arguments.

Let $\widehat{\delta}_1 = \begin{pmatrix} \widehat{\delta}_{y,1}' & \widehat{\delta}_{x,1}' \end{pmatrix}'$, with $\widehat{\delta}_{y,1}$ and $\widehat{\delta}_{x,1}$ initial IV/GMM estimators of δ_y and δ_x in the restricted models (B.1) and (B.2), with $\widehat{\varepsilon}_1 = \begin{pmatrix} \widehat{\varepsilon}_{y,1}' & \widehat{\varepsilon}_{x,1}' \end{pmatrix}'$ the associated residuals. Analogous to the test derived in Section 2, the robust score test statistic for $H_0 : \gamma_y = \gamma_x = 0$ is then given by

$$S_r\left(\widehat{\delta}_1\right) = \widehat{\varepsilon}_1'\widetilde{Z}_2\left(\widetilde{Z}_2'H_{\widehat{\varepsilon}_1}\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'\widehat{\varepsilon}_1, \tag{B.3}$$

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

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

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

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

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

with the Hansen J-test given by

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

Next let $\widehat{\psi}_{1L}$ and $\widehat{\psi}_{2L}$ be the LIML estimators 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 $\widehat{\Sigma}_w^{-1/2} W' P_Z W \widehat{\Sigma}_w^{-1/2}$, and $\widehat{\Sigma}_w = W'W/n$. 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 $\widehat{\varepsilon}_L = \text{vec}\left(W \widehat{\Psi}_L\right)$, where $\widehat{\Psi}_L = \begin{bmatrix} \widehat{\psi}_{1L} & \widehat{\psi}_{2L} \end{bmatrix}$. Then the non-robust score test statistic is given by

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

However, as

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

it follows that

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

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

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

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

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

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

and the resulting estimators $\begin{bmatrix} \widehat{\delta}_{yL} & \widehat{\delta}_{xL} \end{bmatrix} = \widehat{\Psi}_{LB} \widehat{\Psi}_{LA}^{-1}$, see also Cragg and Donald (1993). Clearly, this is the invariant LIML based rank test statistic for testing $H_0 : \text{r}(\Pi^*) = k_x - 1$ 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 statistic is then the robust score test statistic

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

where $\widetilde{Z}_{2L} = I_2 \otimes M_{\widehat{X}_{2L}} Z_o$. 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 statistic is

$$J\left(\widehat{\beta}_{cu}\right) = \min_{\Pi_2,\delta_y,\delta_x} \begin{pmatrix} \widehat{\pi}_y - \Pi_2\delta_y \\ \widehat{\pi}_1 - \Pi_2\delta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix}' \left(V\widehat{ar}_r\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \widehat{\pi}_y - \Pi_2\delta_y \\ \widehat{\pi}_1 - \Pi_2\delta_x \\ \text{vec}\left(\widehat{\Pi}_2 - \Pi_2\right) \end{pmatrix}. \qquad \text{(B.4)}$$

These test statistics are versions of the AHS I test statistic for underidentification for the standard linear IV model. They are straightforwardly generalised to testing for general $H_0 : \text{r}\left(\Pi^*\right) = q$ against $H_1 : \text{r}\left(\Pi^*\right) > q$, the resulting tests being score type tests of the form (B.3), with the LIML and CU-GMM versions being invariant to normalisation.

B.1 Incorporating Cross-Equation Restrictions

One of the motivations of AHS for their I test was that it could incorporate cross-equation restrictions. However, the robust CD rank test as formulated in (B.4) can incorporate cross-equation restrictions. It is 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 $w_t = \left(y_{1,t}\ y_{2,t}\ y_{3,t}\ p_{1,t}\ p_{2,t}\ p_{3,t}\right)'$. 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 $E\left(z_t v_{j,t}\right) = 0$ 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 $H_0 : E\left(z_t v_{j,t}\right) = 0$ for $j = 1, 2, 3$ is in this case a test for $H_0 : \text{r}\left(\Pi^*\right) = 3$, which, incorporating the restrictions, can for example be obtained as the CU-GMM criterion

$$J\left(\widehat{B}_{cu}\right) = \min_{\Pi_p,B} \begin{pmatrix} \text{vec}\left(\widehat{\Pi}_y - \Pi_p B\right) \\ \text{vec}\left(\widehat{\Pi}_p - \Pi_p\right) \end{pmatrix}' \left(V\widehat{ar}_r\left(\widehat{\pi}^*\right)\right)^{-1} \begin{pmatrix} \text{vec}\left(\widehat{\Pi}_y - \Pi_p B\right) \\ \text{vec}\left(\widehat{\Pi}_p - \Pi_p\right) \end{pmatrix},$$

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 for $H_0 : \mathrm{r}(\Pi^*) = 2$. 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 statistic for testing $H_0 : \mathrm{r}(\Pi^*) = 2$, incorporating all restrictions, is given by

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

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 these parameter restrictions 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 $H_1 : \mathrm{r}(\Pi_p) = 3$.

B.2 I Test for Panel AR(2) Model

Arellano, Hansen, and Sentana (1999, 2012) considered the I test in dynamic panel data models, in particular an AR(2) specification in first differences

$$\Delta y_{it} = \alpha_1 \Delta y_{i,t-1} + \alpha_2 \Delta y_{i,t-2} + \Delta u_{it},$$

with moment conditions $E\left[y_i^{t-2}\Delta u_{it}\right] = 0$, where $y_i^{t-2} = \left(y_{i1}, \ldots, y_{i,t-2}\right)'$. Notice that $\Delta y_{i,t-2}$ is here an exogenous variable, perfectly predicted by the instruments, and the standard test for underidentification is like the standard robust F-test for $H_0 : \pi = 0$ in the first-stage equation for the single endogenous variable,

$$\Delta y_{i,-1} = \gamma_1 \Delta y_{i,-2} + Z_i^* \pi + \varepsilon_i,$$

where $\Delta y_{i,-j}$ is the $T-3$ vector $(y_{i,t-j})$, and Z_i^* is the matrix of sequential instruments with one column removed. The robust score test is then a test for overidentification of the $(T-1)(T-2)/2-1$ moment conditions

$$E\left[y_i^{t-2}\left(\Delta y_{i,t-1} - \gamma_1 \Delta y_{i,t-2}\right)\right] = 0, \tag{B.5}$$

for $t = 4, \ldots, T$. Under the null of underidentification, $\pi = 0$, the test statistic will have an asymptotic χ^2 distribution with $(T-1)(T-2)/2-2$ degrees of freedom.

The I test considers testing for underidentification in the sense that there are multiple solutions to the moment conditions $E\left[y_i^{t-2}\Delta u_{it}\right] = 0$, which would be the case if the autoregressive process contains a unit root, such that $\alpha_1 + \alpha_2 = 1$. This is equivalent to testing the null hypothesis $H_0 : \mathrm{r}\left[\pi_y \ \pi\right] = 0$, where π_y are the reduced form parameters in

$$\Delta y_i = \gamma_2 \Delta y_{i,-2} + Z_i^* \pi_y + \nu_i.$$

Because of redundancies due to the autoregressive nature of the process, see Arellano et al. (1999, p 16), the I test is a test for overidentification of the $(T-1)T/2-1$ moment conditions

$$E\left[y_i^{t-2}\left(\Delta y_{i,t-1} - \gamma_1 \Delta y_{i,t-2}\right)\right] = 0, \tag{B.6}$$

for $t = 4, \ldots, T+1$. Compared to moment conditions (B.5), the I test adds the $T-2$ moment conditions $E\left[y_i^{T-1}\left(\Delta y_{i,T} - \gamma_1 \Delta y_{i,T-1}\right)\right] = 0$. These are not in (B.5) because $\Delta y_{i,T}$ is not part of the endogenous explanatory variable.

C Asset Pricing Model Example from Manresa et al. (2017)

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

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

where r_t are excess returns. MPS use the Arellano, Hansen, and Sentana (2012) approach to estimate a basis of the subspace of SDFs compatible with the pricing conditions, instead of a unique SDF (up to scale). Apart from testing the model itself, this basis estimation allows to test important properties of all SDFs in that set, such as for some pricing factor to not enter any SDF, or if all SDFs are uncorrelated with the cross-section of returns.

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 proceed by estimating a basis of that subspace by replicating d times the moment conditions:

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

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

$$E\begin{bmatrix} r_t\left(1 - f_{12,t}'\delta_1\right) \\ r_t\left(1 - f_{13,t}'\delta_2\right) \end{bmatrix} = 0, \tag{C.1}$$

where

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

MPS proceed to estimate the parameters δ_1 and δ_2 by CU-GMM and obtain $J\left(\widehat{\delta}_{cu}\right)$ as a test statistic for the validity of the moment conditions (C.1). Because of invariance, the same test statistic is obtained from specifying the moment conditions as

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

where $f_{23,t}' = \begin{pmatrix} f_{2t} & f_{3t} \end{pmatrix}$. Therefore the $J\left(\widehat{\delta}_{cu}\right)$ test statistic is the same as the robust CD test statistic for testing $H_0 : \mathrm{r}\left(C\right) = k_x - 2$ in

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

following the exposition in Section 3, and is the general robust, CU-GMM based rank test statistic as described in Appendix B.

Similar to the GKR specifications in Section 3.1, the MPS specification is underidentified if $\mathrm{r}(Cov(r_t, f_t)) < k_f$, but an SDF is still economically meaningful if $\mathrm{r}(Cov(r_t x_t')) = \mathrm{r}(Cov(r_t, f_t)) > 0$. The robust CD test statistic for testing $H_0 : Cov(r_t, f_t) = q$ is obtained as the test statistic for testing $H_0 : \mathrm{r}(D) = q$ in

$$f_t = d_0 + D' r_t + v_t.$$

The RANKTEST module for Stata (Kleibergen, Schaffer, and Windmeijer, 2020) reproduces the test and estimation results as reported in e.g. Table 1 of MPS, where they analyse the linear version of the model in Yogo (2006), with the SDF depending on three factors: the market return, and the consumption growth of nondurables and durables. Table C1 presents the results of the robust rank test statistics for a rank reduction of 2 and 3 for the matrices $E(r_t x_t')$ and $Cov(r_t, f_t)$. The results indicate the same rank reduction of 2 for the two matrices, suggesting that the reduction in rank of $E(r_t x_t')$ is due to the reduction in rank of $Cov(r_t, f_t)$ and, as Manresa et al. (2017, p 16) conclude, this suggests that the model is completely overspecified.

Table C1: P-values of robust rank test statistics for $E(r_t x_t')$ and $Cov(r_t, f_t)$

	$E(r_t x_t')$			$Cov(r_t, f_t)$		
rank reduction	KP	J_{2L}	CD	KP	J_{2L}	CD
2	0.107	0.129	0.134	0.113	0.133	0.151
3	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Model and data from Manresa et al. (2017), Table 1.

D Limiting Distribution of the Sargan Test in Underidentified Models

The limiting distributions of the Sargan tests for overidentifying restrictions in the structural model (1), $\mathrm{Sar}\left(\widehat{\beta}_{2sls}\right)$ and $\mathrm{Sar}\left(\widehat{\beta}_L\right)$, when $\mathrm{r}(z_i w_i')$, or equivalently $\mathrm{r}(\Pi^*)$, is less then k_x have been derived by Kitamura (2006) for $\mathrm{Sar}\left(\widehat{\beta}_{2sls}\right)$, and the result of Gospodinov, Kan, and Robotti (2017)) (GKR) derived in the context of linear factor models applies to $\mathrm{Sar}\left(\widehat{\beta}_L\right)$. As $\mathrm{Sar}\left(\widehat{\beta}_L\right)$ is an invariant rank test, its limiting distribution is determined by $\mathrm{r}(\Pi^*)$ only, independent of whether the moments restrictions $E(z_i u_i) = 0$ hold or not. In contrast, the limiting distribution of Sar $\left(\widehat{\beta}_{2sls}\right)$ under rank deficiency depends on whether the moment restrictions hold or not. These limiting distribution results hold under standard maintained assumptions and conditional homoskedasticity.

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 Sar $\left(\widehat{\beta}_L\right)$. Let $\mathrm{r}(\Pi^*) = k_x + 1 - d$, for an integer d. Then the result is that for $d \geq 1$,

$$\text{Sar}\left(\widehat{\beta}_L\right) \xrightarrow{d} w_d, \tag{D.1}$$

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

When the moments $E\left(z_i u_i\right) = 0$ are valid, , the result for Sar $\left(\widehat{\beta}_{2sls}\right)$ as given in Theorem 3.1 in Kitamura (2006, p 67) is,

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

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

When the moment conditions are invalid, the result for Sar $\left(\widehat{\beta}_{2sls}\right)$ as given in Theorem 3.2 in Kitamura (2006, p 71) is,

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

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

Figure D1 displays the limiting distributions of Sar $\left(\widehat{\beta}_L\right)$ and Sar $\left(\widehat{\beta}_{2sls}\right)$ for $k_z - k_x = 7$, for Sar $\left(\widehat{\beta}_{2sls}\right)$ when the moments $E\left(z_i u_i\right) = 0$ are valid, for values of $d = 1, 2, 3$. Figure D2 presents the limiting distribution of Sar $\left(\widehat{\beta}_{2sls}\right)$ 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 Sar $\left(\widehat{\beta}_L\right)$ and Sar $\left(\widehat{\beta}_{2sls}\right)$ is less than nominal size when the moments $E\left(z_i u_i\right) = 0$ are valid, with the discrepancy larger for Sar $\left(\widehat{\beta}_L\right)$ than for Sar $\left(\widehat{\beta}_{2sls}\right)$. Also, as Sar $\left(\widehat{\beta}_L\right)$ 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 $\mathrm{r}(\Pi^*) = k_x$.

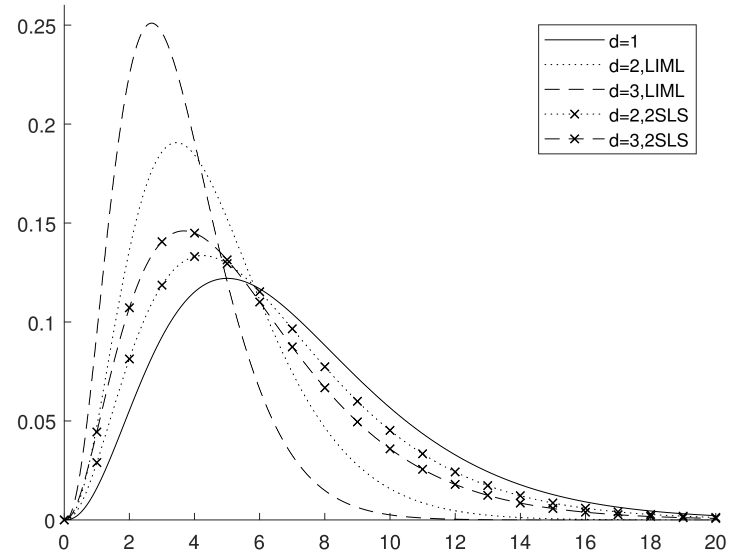

Figure D1: Limiting distributions of Sar $\left(\widehat{\beta}_L\right)$ and Sar $\left(\widehat{\beta}_{2sls}\right)$, for $\mathrm{r}\left(\Pi^*\right) = k_x + 1 - d$, $k_z - k_x = 7$. Moment conditions $E\left(z_i u_i\right) = 0$ are valid for Sar $\left(\widehat{\beta}_{2sls}\right)$.

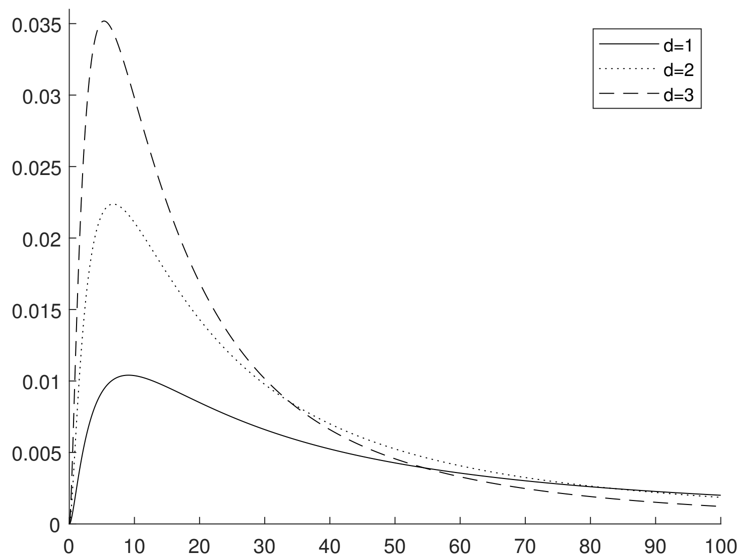

Figure D2: Limiting distributions of Sar $\left(\widehat{\beta}_{2sls}\right)$ for $\mathrm{r}\left(\Pi^*\right) = k_x + 1 - d$, $k_z - k_x = 7$. Moment conditions $E\left(z_i u_i\right) = 0$ are invalid.

In contrast, Figure D2 shows that when the moment conditions do not hold, the limiting distribution of Sar $\left(\widehat{\beta}_{2sls}\right)$ 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.

Table D1 presents the power of Sar $\left(\widehat{\beta}_{2sls}\right)$ 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 (2006, p 74) shows, the limiting distribution of Sar $\left(\widehat{\beta}_{2sls}\right)$ 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 3.1, the degree of overidentification is around 20.

Table D1: Rejection probabilities, $P\left(\text{Sar}\left(\widehat{\beta}_{2sls}\right) > \chi^2_{k_z-k_x,0.95}\right)$

	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 $\chi^2_{k_z-k_x}\chi^2_{k_z-k_x+1}/\chi^2_d$

It is clear from the above results that the invariant Sar $\left(\widehat{\beta}_L\right)$ does not have power to reject $H_0 : E(z_i u_i) = 0$ in underidentified models when the moment conditions are invalid. This is clearly problematic when using Sar $\left(\widehat{\beta}_L\right)$ as a test for overidentifying restrictions. However, this is of course less of a problem for the use of Sar $\left(\widehat{\delta}_L\right)$ as a test for underidentification, and is also a reason why an underidentification test should be reported. The underidentification tests described in Section 2 are for testing $H_0 : \text{r}(\Pi) = k_x - 1$ against $H_1 : \text{r}(\Pi) > k_x - 1$. If $\text{r}(\Pi) < k_x - 1$, the limiting distribution results given in (D.1) apply to Sar $\left(\widehat{\delta}_L\right)$ and so the rejection frequency of Sar $\left(\widehat{\delta}_L\right)$ 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.

E Proofs

Result 1

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

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

with $D = \left[\begin{array}{cc} X & Z_o \end{array} \right]$ and $\theta = \left(\begin{array}{cc} \beta' & \gamma' \end{array} \right)'$. The full instrument matrix is $Z = \left[\begin{array}{cc} Z_1 & Z_o \end{array} \right]$. The null hypothesis can therefore be written as $H_0 : R\theta = 0$, with $R = \left[\begin{array}{cc} O_{k_x} & I_{k_z - k_x} \end{array} \right]$, 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

$$\begin{aligned} \widehat{\theta} &= \left(Z'D\right)^{-1} Z'y \\ V\widehat{a}r\left(\widehat{\theta}\right) &= \left(D'Z\left(Z'H_{\widehat{u}}Z\right)^{-1} Z'D\right)^{-1} = \left(Z'D\right)^{-1} Z'H_{\widehat{u}}Z\left(D'Z\right)^{-1}, \end{aligned}$$

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

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

As

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

where

$$\widehat{D} = P_Z D = \left[\begin{array}{cc} P_Z X & Z_o \end{array} \right] = \left[\begin{array}{cc} \widehat{X} & Z_o \end{array} \right],$$

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

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

it follows that

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

But

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

and so we obtain

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

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

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

As

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

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 $X'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'\widehat{u}_2=0$ and hence $L'D'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'\widehat{u}_2=0$, where $L'=\left[\begin{array}{cc}I_{k_x} & O_{k_z-k_x}\end{array}\right]$. As

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

and letting $B=D'Z\left(Z'H_{\widehat{u}_1}Z\right)^{-1}Z'D$, we get

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

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

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

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

Proposition 1

Proof. 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 $\widehat{\Pi}$. Further, let $Z = \left[\begin{array}{cc} Z_1 & Z_o \end{array}\right]$, with Z_1 an $n \times (k_x - 1)$ matrix, and Π_2 partitioned accordingly as $\Pi_2 = \left[\begin{array}{cc} \Pi_{21}' & \Pi_{2o}' \end{array}\right]'$. It then follows from the discussion in Kleibergen and Paap (2006, pp 101-102), that for $q = k_x - 1$ and $\pi_1 = \Pi_2 \delta$

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

where $\phi = \left(\begin{array}{cc} 1 & -\delta' \end{array}\right)'$. For the test statistic, we can ignore the standardisation terms $\left(I_{k_z-k_x} + \Pi_{2o}\Pi_{21}^{-1}\left(\Pi_{21}'\right)^{-1}\Pi_{2o}'\right)^{-1/2}$ and $\left(\phi'\phi\right)^{-1/2}$. It then follows that the test statistic is based on

$$\widehat{\Lambda}_q = \left[\begin{array}{cc} -\widetilde{\Pi}_{2o}\widetilde{\Pi}_{21}^{-1} & I_{k_z-k_x+1} \end{array}\right]\widehat{\Pi}\widetilde{\phi}, \tag{E.1}$$

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

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

For this case where $G = I_{k_z}$ and $F = I_{k_x}$, it follows that $\widetilde{\Pi}_2$ and $\widetilde{\delta}$ are given by

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

where $\pi_2 = \operatorname{vec}(\Pi_2)$, $\widehat{\pi}_2 = \operatorname{vec}\left(\widehat{\Pi}_2\right)$. Exactly the same formula for $\widehat{\Lambda}_q$ as in (E.1) 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

$$\left(G'G\right)^{-1}G'\left(\widehat{A}_q\widehat{B}_q\right)F\left(F'F\right)^{-1} = \left[\begin{array}{cc} \widetilde{\Pi}_{2,GF} & \widetilde{\Pi}_{2,GF}\widetilde{\delta}_{GF} \end{array}\right],$$

with

$$\left(\widetilde{\Pi}_{2,GF},\widetilde{\delta}_{GF}\right)=\arg\min_{\delta,\Pi_2}\left(\left(\begin{array}{c}\widehat{\pi}_1\\ \widehat{\pi}_2\end{array}\right)-\left(\begin{array}{c}\Pi_2\delta\\ \pi_2\end{array}\right)\right)'(F'F\otimes G'G)\left(\left(\begin{array}{c}\widehat{\pi}_1\\ \widehat{\pi}_2\end{array}\right)-\left(\begin{array}{c}\Pi_2\delta\\ \pi_2\end{array}\right)\right).$$

■

Lemma A.1

Proof. Consider the following minimisation problem

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

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

where 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

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

and the first-order conditions are given by

$$\begin{aligned}\frac{\partial L(\pi^*,\beta,\mu)}{\partial\beta}&=\widetilde{\Pi}'\widetilde{\mu}=0\\ \frac{\partial L(\pi^*,\beta,\mu)}{\partial\mu}&=\widetilde{\Pi}^*\widetilde{\psi}=0,\end{aligned}$$

and

$$\frac{\partial L(\pi^*,\beta,\mu)}{\partial\pi^*}=-(I_{k_x+1}\otimes B)\,A\,(I_{k_x+1}\otimes B)(\widehat{\pi}^*-\widetilde{\pi}^*)+\left(\widetilde{\psi}\otimes\widetilde{\mu}\right)=0.\qquad\text{(E.2)}$$

From (E.2) it follows that

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

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

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

As

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

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

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

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 then 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

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

■

References

Arellano, M., L. P. Hansen, and E. Sentana (1999): "Underidentification?" Mimeo, CEMFI, ftp://ftp.cemfi.es/pdf/papers/es/Ahsc.pdf.

——— (2012): "Underidentification?" *Journal of Econometrics*, 170, 256–280.

Cragg, J. G. and S. G. Donald (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

Gospodinov, N., R. Kan, and C. Robotti (2017): "Spurious Inference in Reduced-Rank Asset-Pricing Models," *Econometrica*, 85, 1613–1628.

Kitamura, Y. (2006): "Specification Tests with Instrumental Variables and Rank Deficiency," in *Econometric Theory and Practice*, ed. by D. Corbae, S. N. Durlauf, and B. E. Hansen, Cambridge University Press, 59–81.

Kleibergen, F. and R. Paap (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

Kleibergen, F., M. E. Schaffer, and F. Windmeijer (2020): "RANKTEST: Stata Module to Test the Rank of a Matrix," https://ideas.repec.org/c/boc/bocode/s456865.html.

Manresa, E., F. Peñaranda, and E. Sentana (2017): "Empirical Evaluation of Overspecified Asset Pricing Models," CEPR Discussion Paper Series DP12085, CEPR, London.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge and London, second ed.

Yogo, M. (2006): "A Consumption-Based Explanation of Expected Stock Returns," *The Journal of Finance*, 61, 539–580.