Two-Stage Least Squares as Minimum Distance

Frank Windmeijer

Discussion Paper 17 / 683

7 June 2017

Revised 13 June 2018

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

Two-Stage Least Squares as Minimum Distance

Frank Windmeijer*
Department of Economics and IEU
University of Bristol, UK

First version June 2017. This version May 2018.

Abstract

The Two-Stage Least Squares instrumental variables (IV) estimator for the parameters in linear models with a single endogenous variable is shown to be identical to an optimal Minimum Distance (MD) estimator based on the individual instrument specific IV estimators. The 2SLS estimator is a linear combination of the individual estimators, with the weights determined by their variances and covariances under conditional homoskedasticity. It is further shown that the Sargan test statistic for overidentifying restrictions is the same as the MD criterion test statistic. This provides an intuitive interpretation of the Sargan test. The equivalence results also apply to the efficient two-step GMM and robust optimal MD estimators and criterion functions, allowing for general forms of heteroskedasticity. It is further shown how these results extend to the linear overidentified IV model with multiple endogenous variables.

JEL Classification: C26, C13, C12
Key Words: Instrumental Variables, Two-Stage Least Squares, Minimum Distance, Overidentification Test

*f.windmeijer@bristol.ac.uk. This research was partly funded by the Medical Research Council (MC_UU_12013/9). I would like to thank a referee, Victor Chernozhukov, Steve Pischke, Mark Schaffer, Chris Skeels and Jon Temple for helpful comments.

1 Introduction

For a single endogenous variable linear model with multiple instruments, the standard IV estimator is the Two-Stage Least Squares (2SLS) estimator, which is a consistent and asymptotically efficient estimator under standard regularity assumptions and conditional homoskedasticity, see e.g. Hayashi (2000, p 228). This means that the 2SLS estimator combines the information from the multiple instruments asymptotically optimally under these conditions. An alternative estimator is the optimal Minimum Distance (MD) estimator, using an estimator of the variance matrix of the individual instrument-specific IV estimators of the parameter of interest. It is shown in the next section, Section 2, that this optimal MD estimator, with the variance specified under conditional homoskedasticity, is identical to the 2SLS estimator. It is further shown that the Sargan test statistic for overidentifying restrictions is the same as the MD criterion test statistic, providing another intuitive interpretation of the Sargan test.

Surprisingly, it appears that these equivalence results are not available in the literature, and are not discussed in standard textbooks. Angrist (1991) derives similar results, but for the special case of orthogonal binary instruments, see also the discussion in Angrist and Pischke (2009, Section 4.2.2), whereas the results here are for general designs. Recently, Chen, Jacho-Chávez and Linton (2016) used this setting and the two estimators as an example in their much wider-ranging paper, but they did not realise their equivalence and the results obtained in Section 2 modify the statements of Chen et al. (2016, pp 48-49).

In Section 2.2, the result is extended to the equivalence of the two-step GMM estimator and the optimal minimum distance estimator based on a robust variance-covariance estimator of the vector of instrument-specific IV estimates, robust to general forms of heteroskedasticity in the cross-sectional setting considered here. The two-step Hansen J-test statistic for overidentifying restrictions (Hansen, 1982) is also shown to be the same as the robust MD criterion test statistic.

Section 3 derives equivalence results for the multiple endogenous variables case. The setting considered there can best be characterised by the following simple example. Consider a linear model with two endogenous variables, and there are four instruments available. In principle, there are then six distinct sets of two, just identifying instruments.

However, a collection of three sets of two instruments that span all instruments is sufficient to provide all information needed. For example, if the instruments are denoted by z_1, z_2, z_3, and z_4, then the collection of sets $\{(z_1, z_2), (z_2, z_3), (z_3, z_4)\}$ is sufficient. This results in three just identified IV estimates of the two parameters of interest, and Section 3 shows that the per parameter optimal minimum distance estimators are identical to the 2SLS estimators.

2 Equivalence Result for Single Endogenous Variable Model

We have a sample $\{(y_i, x_i, z_i')\}_{i=1}^n$ and consider the model

$$y_i = x_i\beta + u_i,$$

where x_i is endogenous, such that $E(x_i u_i) \neq 0$. Note that other exogenous variables in the model, including the constant, have been partialled out. The $k_z > 1$ instrument vector z_i satisfies $E(z_i u_i) = 0$ and is related to x_i via the linear projection, or first-stage model

$$x_i = z_i'\pi_x + v_i. \tag{1}$$

Let y and x be the n-vectors $(y_1, y_2, ..., y_n)'$ and $(x_1, x_2, ..., x_n)'$, and Z the $n \times k_z$ matrix with i-th row z_i' and j-th column z_j, $i = 1, .., n$, $j = 1, ..., k_z$.

Let $P_Z = Z(Z'Z)^{-1}Z'$ and

$$S(b) = (y - xb)' P_Z (y - xb). \tag{2}$$

The well-known Two-Stage Least Squares (2SLS) instrumental variables estimator is then defined as

$$\widehat{\beta}_{2SLS} = \arg\min_b S(b)$$

and is given by

$$\widehat{\beta}_{2SLS} = (x'P_Z x)^{-1} x'P_Z y. \tag{3}$$

Next consider the individual instrument-specific IV estimators for β, given by

$$\widehat{\beta}_j = (z_j'x)^{-1} z_j'y,$$

for $j = 1, .., k_z$. Let the k_z-vector $\widehat{\beta}_{ind}$ be defined as

$$\widehat{\beta}_{ind} = \left(\widehat{\beta}_1, \widehat{\beta}_2, ..., \widehat{\beta}_{k_z}\right)', \tag{4}$$

then

$$\widehat{\beta}_{ind} = D_{zx}^{-1} Z' y,$$

where $D_{zx} = \text{diag}\left(z_j' x\right)$. The matrix $\text{diag}\left(a_j\right)$ is a diagonal matrix with j-th diagonal element a_j.

It follows that the 2SLS estimator is a linear combination of the individual estimators, as

$$\begin{aligned} \widehat{\beta}_{2SLS} &= \left(x' P_Z x\right)^{-1} x' Z \left(Z'Z\right)^{-1} D_{zx} \widehat{\beta}_{ind} \\ &= \sum_{j=1}^{k_z} w_{2sls,j} \widehat{\beta}_j, \end{aligned} \tag{5}$$

with $\sum_{j=1}^{k_z} w_{2sls,j} = 1$.

Under the assumptions that $\frac{1}{n}\sum_{i=1}^{n} z_j' x \xrightarrow{p} \lambda_j \neq 0$ for $j = 1, ..., k_z$, $\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \xrightarrow{p} H_{zz}$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i u_i \xrightarrow{d} N\left(0, \sigma_u^2 H_{zz}\right)$, the limiting distribution of $\widehat{\beta}_{ind}$ is given by

$$\sqrt{n}\left(\widehat{\beta}_{ind} - \iota\beta\right) = \sqrt{n} D_{zx}^{-1} Z' u \xrightarrow{d} N\left(0, \sigma_u^2 \Omega\right), \tag{6}$$

where ι is a k_z-vector of ones, and Ω is given by

$$\Omega = D_\lambda^{-1} H_{zz} D_\lambda^{-1},$$

with $D_\lambda = \text{diag}\left(\lambda_j\right)$.

Let

$$\widehat{\Omega} = D_{zx}^{-1} Z' Z D_{zx}^{-1},$$

with $n\widehat{\Omega} \xrightarrow{p} \Omega$. The optimal minimum distance (MD) estimator for β is then given by

$$\begin{aligned} \widehat{\beta}_{md} &= \arg\min_b Q\left(b\right); \\ Q\left(b\right) &= \left(\widehat{\beta}_{ind} - \iota b\right)' \widehat{\Omega}^{-1} \left(\widehat{\beta}_{ind} - \iota b\right), \end{aligned} \tag{7}$$

resulting in

$$\widehat{\beta}_{md} = \left(\iota' \widehat{\Omega}^{-1} \iota\right)^{-1} \iota' \widehat{\Omega}^{-1} \widehat{\beta}_{ind}. \tag{8}$$

It is clear that the the MD estimator is also a linear combination of the individual instrument specific estimators,

$$\widehat{\beta}_{md} = \sum_{j=1}^{k_z} w_{md,j} \widehat{\beta}_j,$$

with $\sum_{j=1}^{k_z} w_{md,j} = 1$.The next proposition states the main equivalence result, $Q(b) = S(b)$, hence $\widehat{\beta}_{md} = \widehat{\beta}_{2sls}$ and $w_{md,j} = w_{2sls,j}$ for $j = 1, .., k_z$.

Proposition 1 *Let $S(b)$, $Q(b)$, $\widehat{\beta}_{2sls}$ and $\widehat{\beta}_{md}$ be as defined in (2), (7), (3) and (8) respectively. Then for $b \in \mathbb{R}$, $Q(b) = S(b)$ and hence $\widehat{\beta}_{md} = \widehat{\beta}_{2sls}$.*

Proof. As $\iota' D_{zx} = Z'x$, it follows that, for $b \in \mathbb{R}$,

$$\begin{aligned} Q(b) &= \left(\widehat{\beta}_{ind} - \iota b\right)' \widehat{\Omega}^{-1} \left(\widehat{\beta}_{ind} - \iota b\right) \\ &= \left(D_{zx}^{-1} Z'y - \iota b\right)' D_{zx} \left(Z'Z\right)^{-1} D_{zx} \left(D_{zx}^{-1} Z'y - \iota b\right) \\ &= (y - xb)' Z \left(Z'Z\right)^{-1} Z' (y - xb) \\ &= S(b). \end{aligned}$$

■

Note that the equivalence results obtained in Proposition 1 does, given the choice of $\widehat{\Omega}$, not rely on any high level assumptions. For example, whilst the limiting distribution of $\widehat{\beta}_{ind}$ in (6) can only be derived under the assumption that $\lambda_j \neq 0$, for all $j = 1, ..k_z$, the numerical equivalence results hold also when this assumption is violated and even if $\lambda_j = 0$ for all j.

Let $w_j = w_{md,j} = w_{2sls,j}$. Whilst $\sum_{j=1}^{k_z} w_j = 1$, the weigths w_j can be negative, in which case $\widehat{\beta}_{2sls}$ is not a weighted average of the $\widehat{\beta}_j$. From the definition of w_j in (5) it follows that $\text{sign}(w_j) = \text{sign}\left(\widehat{\pi}_{x,j} \left(z_j' x\right)\right)$, where $\widehat{\pi}_{x,j}$ is the j-th element of the OLS estimator of π_x in (1). Wlog, we can code the instruments such that $\widehat{\pi}_{x,j} \geq 0$ for all j, and standardise such that $z_j' z_j / n = 1$. It then follows that $\text{sign}(w_j) = \text{sign}\left(\widehat{\lambda}_j\right)$, where $\widehat{\lambda}_j$ is the OLS estimator of λ_j in the first-stage specification $x = z_j \lambda_j + v_j$. Therefore, $w_j \geq 0$ for all j iff $\widehat{\lambda}_j \geq 0$ for all j. As $\widehat{\lambda}_j = \left(z_j' Z / n\right) \widehat{\pi}_x = \widehat{\pi}_{x,j} + \sum_{l=1, l \neq j}^{k_z} \widehat{\rho}_{jl} \widehat{\pi}_{x,l}$, where $\widehat{\rho}_{jl} = z_j' z_l / n = \widehat{\rho}_{lj}$, it follows that $\widehat{\lambda}_j \geq 0$ if $\sum_{l=1, l \neq j}^{k_z} \widehat{\rho}_{jl} \widehat{\pi}_{x,l} \geq -\widehat{\pi}_{x,j}$. A sufficient condition

for $w_j \geq 0$ for all j is then that $\widehat{\rho}_{jl} \geq 0$ for all j,l, $l > j$, i.e. the instruments are uncorrelated or positively correlated with each other.

The weights for the minimum distance estimator are obtained from the constrained minimisation problem

$$w_{md} = \arg\min_{w} w'\widehat{\Omega}w \quad \text{s.t.} \quad \sum_{j=1}^{k_z} w_j = 1.$$

Imposing the constraints $w_j \geq 0$, for $j = 1, ..., k_z$, results in a standard quadratic programming problem. If there are negative weigths in the original solution, then imposing nonnegativity will lead to some of the $w_{md,j}$ set equal to zero. The resulting estimator is then equal to a weighted average of the $\widehat{\beta}_j$ for a subset of the instruments that minimises the variance over the subsets for which the 2SLS and MD estimators are weighted averages of the instrument specific estimates.

2.1 Test for Overidentifying Restrictions

The standard test for the null hypothesis $H_0 : E(z_i u_i) = 0$ is the Sargan test statistic given by

$$Sar\left(\widehat{\beta}_{2sls}\right) = \widehat{\sigma}_u^{-2} S\left(\widehat{\beta}_{2sls}\right),$$

where $\widehat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i\widehat{\beta}_{2sls}\right)^2$. Under the null, standard regularity assumptions and conditional homoskedasticity, $Sar\left(\widehat{\beta}_{2sls}\right)$ converges in distribution to a $\chi^2_{k_z-1}$ distributed random variable, see e.g. Hayashi (2000, p 228).

Next consider the MD criterion

$$MD\left(\widehat{\beta}_{md}\right) = \widehat{\sigma}_u^{-2} Q\left(\widehat{\beta}_{md}\right),$$

where we can use $\widehat{\sigma}_u^2$ because $\widehat{\beta}_{md} = \widehat{\beta}_{2sls}$. Let $\beta_j = \text{plim}\left(\widehat{\beta}_j\right)$. Under the null hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_{k_z} = \beta$, and the assumptions stated above for the limiting distribution (6) to hold, $MD\left(\widehat{\beta}_{md}\right)$ converges in distribution to a $\chi^2_{k_z-1}$ distributed random variable, see e.g. Cameron and Trivedi (2005, p 203).

It follows directly from the results of Proposition 1 that $S\left(\widehat{\beta}_{2sls}\right) = Q\left(\widehat{\beta}_{md}\right)$ and hence

$$Sar\left(\widehat{\beta}_{2sls}\right) = MD\left(\widehat{\beta}_{md}\right).$$

Remark 1 *The equivalence of* $Sar\left(\widehat{\beta}_{2sls}\right)$ *and* $MD\left(\widehat{\beta}_{md}\right)$ *establishes an intuitive interpretation of the Sargan test. It tests whether the individual instrument-specific estimators all estimate the same parameter value. For a related discussion, see Parente and Santos Silva (2012).*

2.2 Efficient Two-Step Estimation

The equivalence results extend to the efficient two-step GMM estimator. For the cross-sectional setup considered here, this would cover the case of general conditional heteroskedasticity, $E\left(u_i^2|z_i\right) = g\left(z_i\right)$. Assume $\frac{1}{n}\sum_{i=1}^{n} u_i^2 z_i z_i' \xrightarrow{p} \Sigma$. Using $\widehat{\beta}_{2sls}$ as the initial consistent one-step GMM estimator, the efficient two-step GMM estimator is defined as

$$\widehat{\beta}_{gmm} = \arg\min_{b} J\left(b\right),$$

where

$$\begin{aligned} J\left(b\right) &= \left(y - xb\right)' Z\widehat{\Sigma}^{-1}\left(\widehat{\beta}_{2sls}\right) Z'\left(y - xb\right); \\ \widehat{\Sigma}\left(\widehat{\beta}_{2sls}\right) &= \sum_{i=1}^{n}\left(y_i - x_i\widehat{\beta}_{2sls}\right)^2 z_i z_i'. \end{aligned}$$

The Hansen J-test for overidentifying restrictions is given by $J\left(\widehat{\beta}_{gmm}\right)$. Under standard assumptions, $J\left(\widehat{\beta}_{gmm}\right) \xrightarrow{d} \chi^2_{k_z - 1}$ under the null $H_0 : E\left(z_i u_i\right) = 0$.

Under the assumptions as stated above, the limiting distribution of $\widehat{\beta}_{ind}$ is given by

$$\sqrt{n}\left(\widehat{\beta}_{ind} - \iota\beta\right) \xrightarrow{d} N\left(0, \Omega_r\right),$$

where

$$\Omega_r = D_\lambda^{-1}\Sigma D_\lambda^{-1},$$

and, as $\widehat{\beta}_{md} = \widehat{\beta}_{2sls}$, a robust variance estimator for $\widehat{\beta}_{ind}$ is given by

$$\widehat{\Omega}_r = D_{zx}^{-1}\widehat{\Sigma}\left(\widehat{\beta}_{2sls}\right) D_{zx}^{-1},$$

Define the robust MD estimator as

$$\begin{aligned} \widehat{\beta}_{md,r} &= \arg\min_{b} MD_r\left(b\right); \\ MD_r\left(b\right) &= \left(\widehat{\beta}_{ind} - \iota b\right)' \widehat{\Omega}_r^{-1}\left(\widehat{\beta}_{ind} - \iota b\right). \end{aligned}$$

Under the null as specified above, $H_0 : \beta_1 = \beta_2 = ... = \beta_{k_z} = \beta$, and the assumptions stated above, $MD_r\left(\widehat{\beta}_{md,r}\right)$ converges in distribution to a $\chi^2_{k_z-1}$ distributed random variable.

It follows directly from the proof of Proposition 1 that, for $b \in \mathbb{R}$, $MD_r(b) = J(b)$ and hence $\widehat{\beta}_{gmm} = \widehat{\beta}_{md,r}$ and $J\left(\widehat{\beta}_{gmm}\right) = MD_r\left(\widehat{\beta}_{md,r}\right)$.

Remark 2 *An alternative "one-step" robust variance estimator for the MD estimator is given by*

$$\widehat{\Omega}_{r,ind} = D_{zx}^{-1}\widehat{\Sigma}\left(\widehat{\beta}_{ind}\right)D_{zx}^{-1},$$

with the elements of $\widehat{\Sigma}\left(\widehat{\beta}_{ind}\right)$ *given by*

$$\widehat{\Sigma}\left(\widehat{\beta}_{ind}\right)_{j,l} = \sum_{i=1}^{n}\left(y_i - x_i\widehat{\beta}_j\right)\left(y_i - x_i\widehat{\beta}_l\right)z_{ij}z_{il},$$

for $j, l = 1, ..., k_z$. *The resulting minimum distance estimator,* $\widehat{\beta}_{md,ind}$, *has the same limiting distribution as* $\widehat{\beta}_{md,r}$, *but differs in finite samples.*

Note that the minimum distance objective function we consider here is different from the minimum distance approach that leads for example to the LIML and Continuously Updating (CU) GMM estimators. Consider the OLS estimators $\widehat{\pi}_x$ and $\widehat{\pi}_y$ for π_x in model (1) and π_y in the specification $y_i = z_i'\pi_y + \varepsilon_i = z_i'(\beta\pi_x) + u_i + \beta v_i$. Then consider the minimum distance estimator

$$\left(\widehat{\beta}_{nmd}, \widehat{\pi}_{x,nmd}\right) = \arg\min_{\beta,\pi_x}\begin{pmatrix}\widehat{\pi}_y - \beta\pi_x \\ \widehat{\pi}_x - \pi_x\end{pmatrix}'\widehat{V}^{-1}\begin{pmatrix}\widehat{\pi}_y - \beta\pi_x \\ \widehat{\pi}_x - \pi_x\end{pmatrix},$$

where $\widehat{V} = \widehat{Var}\left(\begin{pmatrix}\widehat{\pi}_y' & \widehat{\pi}_x'\end{pmatrix}'\right)$. If $\widehat{V}$ is a valid variance estimator under conditional homoskedasiticy only, $\widehat{\beta}_{nmd}$ is equal to the LIML estimator, see Goldberger and Olkin (1971). If $\widehat{V}$ is a robust variance estimator, $\widehat{\beta}_{nmd}$ is the CU-GMM estimator, see the discussion in Windmeijer (2018). Other recent approaches to minimum distance estimation are Sølvsten (2017) and Kolesár (2018).

3 Multiple Endogenous Variables

Consider next the multiple endogenous variables model

$$y_i = x_i'\beta + u_i,$$

where x_i is a k_x vector of endogenous variables. There are $k_z > k_x$ instruments z_i available. Let X be the $n \times k_x$ matrix of explanatory variables, with l-th column x_l, then the 2SLS estimator is obtained as

$$\begin{aligned} \widehat{\beta}_{2sls} &= \arg\min_b S(b) \\ S(b) &= (y - Xb)' P_Z (y - Xb) \end{aligned}$$

and is given by

$$\widehat{\beta}_{2sls} = (X' P_Z X)^{-1} X' P_Z y.$$

An MD estimator could of course be obtained here in similar fashion to the one-variable case above, from $\binom{k_z}{k_x}$ sets of just-identifying instruments and a generalized inverse for the now rank deficient variance matrix $\widehat{\Omega}$. Calculating the variance matrix under conditional homoskedasticity leads again to equivalence of the 2SLS and MD estimators.

However, more interesting results can be derived for the 2SLS and MD estimators of the individual coefficients β_l, $l = 1, .., k_x$. Denote by $\widehat{X}_{.l}$ the l-th column of $\widehat{X}$, and let $\widehat{X}_{-l}$ be the $k_x - 1$ columns of $\widehat{X}$, excluding $\widehat{X}_{.l}$. The 2SLS estimator for β_l is given by

$$\begin{aligned} \widehat{\beta}_{l,2sls} &= \left(\widehat{X}_{.l}' M_{\widehat{X}_{-l}} \widehat{X}_{.l}\right)^{-1} \widehat{X}_{.l}' M_{\widehat{X}_{-l}} y \\ &= (\widetilde{x}_l' \widetilde{x}_l)^{-1} \widetilde{x}_l' y \end{aligned} \tag{9}$$

where for a general $n \times k$ matrix A, $M_A = I_n - P_A$, with I_n the identity matrix of order n, and

$$\widetilde{x}_l = M_{\widehat{X}_{-l}} \widehat{X}_{.l}. \tag{10}$$

Let $\left\{Z^{[t]}\right\}_{t=1}^{k_z - k_x + 1}$ be a collection of $k_z - k_x + 1$ sets of k_x instruments $Z^{[t]}$ such that all instruments have been included. For example, $\left\{Z^{[t]} = (z_t, ..., z_{t+k_x-1})\right\}_{t=1}^{k_z - k_x + 1}$ is such a set. From these sets, we get $k_z - k_x + 1$ just identified IV estimates $\widehat{\beta}^{[t]}$ of β. Let $\widehat{\beta}_{l,ind} = \left(\widehat{\beta}_l^{[t]}\right)$ be the $(k_z - k_x + 1)$-vector of the individual estimates of β_l. Let

$$\widehat{X}^{[t]} = P_{Z^{[t]}} X,$$

and $\widehat{X}_{.l}^{[t]}$ and $\widehat{X}_{-l}^{[t]}$ defined analogously to above. The elements of $\widehat{\beta}_{l,ind}$ are then given by

$$\begin{aligned} \widehat{\beta}_{l,ind}^{[t]} &= \left(\widehat{X}_{.l}^{[t]\prime} M_{\widehat{X}_{-l}^{[t]}} \widehat{X}_{.l}^{[t]}\right)^{-1} \widehat{X}_{.l}^{[t]\prime} M_{\widehat{X}_{-l}^{[t]}} y \\ &= \left(\widetilde{x}_l^{[t]\prime} \widetilde{x}_l^{[t]}\right)^{-1} \widetilde{x}_l^{[t]\prime} y, \end{aligned}$$

for $t = 1, ..., k_z - k_x + 1$, where

$$\widetilde{x}_l^{[t]} = M_{\widehat{X}_{-l}^{[t]}} \widehat{X}_{.l}^{[t]}. \tag{11}$$

Hence,

$$\widehat{\beta}_{l,ind} = \widetilde{D}_l^{-1} \widetilde{X}_l' y, \tag{12}$$

with

$$\widetilde{X}_l = \left(\widetilde{x}_l^{[1]}, ..., \widetilde{x}_l^{[k_z - k_x + 1]} \right)$$

and $\widetilde{D}_l = \text{diag}\left(\widetilde{x}_l^{[t]\prime} \widetilde{x}_l^{[t]} \right)$, $t = 1, ..., k_z - k_x + 1$. From (12) and the proof of Proposition 2 below, it follows that $\sqrt{n}\left(\widetilde{D}_l^{-1} \widetilde{X}_l' y - \iota\beta \right) = \sqrt{n} \widetilde{D}_l^{-1} \widetilde{X}_l' u$. Hence, the variance of $\widehat{\beta}_{l,ind}$ under conditional homoskedasticity can be specified as

$$\begin{aligned} \text{Var}\left(\widehat{\beta}_{l,ind} \right) &= \sigma_u^2 \widehat{\Omega}_l \\ \widehat{\Omega}_l &= \widetilde{D}_l^{-1} \widetilde{X}_l' \widetilde{X}_l \widetilde{D}_l^{-1}. \end{aligned}$$

The MD estimator for β_l is then obtained as

$$\begin{aligned} \widehat{\beta}_{l,md} &= \arg\min_b Q_l(b); \\ Q_l(b) &= \left(\widehat{\beta}_{l,ind} - \iota b \right)' \widehat{\Omega}_l^{-1} \left(\widehat{\beta}_{l,ind} - \iota b \right) \end{aligned}$$

where ι is here a $k_z - k_x + 1$ vector of ones. $\widehat{\beta}_{l,md}$ is therefore given by

$$\begin{aligned} \widehat{\beta}_{l,md} &= \left(\iota' \widetilde{D}_l \left(\widetilde{X}_l' \widetilde{X}_l \right)^{-1} \widetilde{D}_l \iota \right)^{-1} \iota' \widetilde{D}_l \left(\widetilde{X}_l' \widetilde{X}_l \right)^{-1} \widetilde{D}_l \widehat{\beta}_{l,ind} \\ &= \left(\iota' \widetilde{D}_l \left(\widetilde{X}_l' \widetilde{X}_l \right)^{-1} \widetilde{D}_l \iota \right)^{-1} \iota' \widetilde{D}_l \left(\widetilde{X}_l' \widetilde{X}_l \right)^{-1} \widetilde{X}_l' y. \end{aligned} \tag{13}$$

The next proposition establishes the equivalence of $\widehat{\beta}_{l,2sls}$ and $\widehat{\beta}_{l,md}$ for $l = 1, ..., k_x$.

Proposition 2 *For $l = 1, ..., k_x$, let $\widehat{\beta}_{l,2sls}$, $\widehat{\beta}_{l,ind}$ and $\widehat{\beta}_{l,md}$ be as defined in (9), (12) and (13) respectively, with $\widehat{\beta}_{l,ind}$ based on a collection of $k_z - k_x + 1$ sets of k_x instruments $\left\{ Z^{[t]} \right\}_{t=1}^{k_z - k_x + 1}$ that contains all instruments. Then $\widehat{\beta}_{l,2sls} = \widehat{\beta}_{l,md}$ for $l = 1, ..., k_x$.*

Proof. From the definitions of $\widetilde{x}_l$ and $\widetilde{x}_l^{[t]}$ in (10) and (11) respectively, it follows that

$$\begin{aligned} \widetilde{x}_l' \widetilde{x}_l^{[t]} &= x_l' \left(P_Z - P_Z X_{-l} \left(X_{-l}' P_Z X_{-l} \right)^{-1} X_{-l}' P_Z \right) \\ &\quad * \left(P_{Z^{[t]}} - P_{Z^{[t]}} X_{-l} \left(X_{-l}' P_{Z^{[t]}} X_{-l} \right)^{-1} X_{-l}' P_{Z^{[t]}} \right) x_l \\ &= x_l' \left(P_{Z^{[t]}} - P_{Z^{[t]}} X_{-l} \left(X_{-l}' P_{Z^{[t]}} X_{-l} \right)^{-1} X_{-l}' P_{Z^{[t]}} \right) x_l \\ &= x_l' \widetilde{x}_l^{[t]} = \widetilde{x}_l^{[t]\prime} \widetilde{x}_l^{[t]}, \end{aligned}$$

for $t = 1, ..., k_z - k_x + 1$, and hence

$$\iota' \widetilde{D}_l = \widetilde{x}_l' \widetilde{X}_l.$$

Therefore, from (13),

$$\widehat{\beta}_{l,md} = \left(\widetilde{x}_l' P_{\widetilde{X}_l} \widetilde{x}_l \right)^{-1} \widetilde{x}_l' P_{\widetilde{X}_l} y.$$

As the sets of instruments $\left\{ Z^{[t]} \right\}_{t=1}^{k_z - k_x + 1}$ contain all k_z instruments, it follows that $\widetilde{x}_l$ is in the column space of $\widetilde{X}_l$, and so $P_{\widetilde{X}_l} \widetilde{x}_l = \widetilde{x}_l$. Therefore,

$$\widehat{\beta}_{l,md} = \left(\widetilde{x}_l' \widetilde{x}_l \right)^{-1} \widetilde{x}_l' y = \widehat{\beta}_{l,2sls},$$

for $l = 1, ..., k_x$. ■

Next, consider the Sargan test statistic, given by

$$Sar \left(\widehat{\beta}_{2sls} \right) = \widehat{\sigma}_u^{-2} S \left(\widehat{\beta}_{2sls} \right), \tag{14}$$

where $\widehat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - x_i' \widehat{\beta}_{2sls} \right)^2$. Under the null $H_0 : E \left(z_i u_i \right) = 0$, standard regularity conditions and conditional homoskedasticity, $Sar \left(\widehat{\beta}_{2sls} \right) \xrightarrow{d} \chi^2_{k_z - k_x}$.

Consider the MD statistics

$$MD \left(\widehat{\beta}_{l,md} \right) = \widehat{\sigma}_u^{-2} Q \left(\widehat{\beta}_{l,md} \right), \tag{15}$$

for $l = 1, ..., k_x$. Let $\beta_{l,ind} = \text{plim} \left(\widehat{\beta}_{l,ind} \right)$, then $MD \left(\widehat{\beta}_{l,md} \right) \xrightarrow{d} \chi^2_{k_z - k_x}$ under the null $H_0 : \beta_{l,ind} = \iota \beta_l$, but as $\widehat{\sigma}_u^2$ has to be a consistent estimator of σ_u^2, the maintained assumptions are that $\beta_{s,ind} = \iota \beta_s$ for $s = 1, ..., k_x$, $s \neq l$.

The following proposition states the equivalence of $Sar \left(\widehat{\beta}_{2sls} \right)$ and $MD \left(\widehat{\beta}_{l,md} \right)$ for $l = 1, ..., k_x$.

Proposition 3 *Let* $Sar \left(\widehat{\beta}_{2sls} \right)$ *and* $MD \left(\widehat{\beta}_{l,md} \right)$ *be defined as in (14) and (15), then* $Sar \left(\widehat{\beta}_{2sls} \right) = MD \left(\widehat{\beta}_{l,md} \right)$ *for* $l = 1, ..., k_x$.

Proof. As $\widehat{X}' \left(y - X \widehat{\beta}_{2sls} \right) = 0$, $\widehat{\beta}_{l,2sls} = \widehat{\beta}_{l,md}$, and defining $\widetilde{y} = M_{\widehat{X}_{-l}} P_Z y$, it follows

that

$$\begin{aligned}
S\left(\widehat{\beta}_{2sls}\right) &= \left(y - X\widehat{\beta}_{2sls}\right)' P_Z \left(y - X\widehat{\beta}_{2sls}\right) \\
&= \left(y - x_l\widehat{\beta}_{l,2sls} - X_{-l}\widehat{\beta}_{-l,2sls}\right)' P_Z M_{\widehat{X}_{-l}} P_Z \left(y - x_l\widehat{\beta}_{l,2sls} - X_{-l}\widehat{\beta}_{-l,2sls}\right) \\
&= \left(\widetilde{y} - \widetilde{x}_l\widehat{\beta}_{l,2sls}\right)' \left(\widetilde{y} - \widetilde{x}_l\widehat{\beta}_{l,2sls}\right) \\
&= \left(\widetilde{y} - \widetilde{x}_l\widehat{\beta}_{l,2sls}\right)' P_{\widetilde{X}_l} \left(\widetilde{y} - \widetilde{x}_l\widehat{\beta}_{l,2sls}\right) \\
&= \left(\widetilde{D}_l^{-1}\widetilde{X}_l'\widetilde{y} - \widetilde{D}_l^{-1}\widetilde{X}_l'\widetilde{x}_l\widehat{\beta}_{l,md}\right)' \widetilde{D}_l \left(\widetilde{X}_l'\widetilde{X}_l\right)^{-1} \widetilde{D}_l \left(\widetilde{D}_l^{-1}\widetilde{X}_l'\widetilde{y} - \widetilde{D}_l^{-1}\widetilde{X}_l'\widetilde{x}_l\widehat{\beta}_{l,md}\right) \\
&= \left(\widehat{\beta}_{l,ind} - \iota\widehat{\beta}_{l,md}\right)' \widehat{\Omega}_l^{-1} \left(\widehat{\beta}_{l,ind} - \iota\widehat{\beta}_{l,md}\right) \\
&= Q\left(\widehat{\beta}_{l,md}\right)
\end{aligned}$$

and hence $Sar\left(\widehat{\beta}_{2sls}\right) = MD\left(\widehat{\beta}_{l,md}\right)$ for $l = 1, ..., k_x$. ■

As for the single-endogenous variable case, these results can be extended to the two-step GMM and robust MD estimators.

References

Angrist, J.D., 1991, Grouped-Data Estimation and Testing in Simple Labor-Supply Models, *Journal of Econometrics* 47, 243-266.

Angrist, J.D. and J.-S. Pischke, 2009, *Mostly Harmless Econometrics. An Empiricist's Companion*, Princeton University Press, Princeton and Oxford.

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

Chen, X., D.T. Jacho-Chávez and O. Linton, 2016, Averaging of an Increasing Number of Moment Condition Estimators, *Econometric Theory* 32, 30-70.

Goldberger, A.S. and I. Olkin, 1971, A Minimum-Distance Interpretation of Limited-Information Estimation, *Econometrica* 39, 635-639.

Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029-1054.

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

Kolesár, M., 2018, Minimum Distance Approach to Inference with Many Instruments, *Journal of Econometrics* 204, 86-100.

Sølvsten, M., 2017, Robust Estimation with Many Instruments, mimeo, University of Wisconsin.

Parente, P.M.C. and J.M.C. Santos Silva (2012) A Cautionary Note on Tests of Overidentifying Restrictions, *Economics Letters* 115, 314-317.

Windmeijer, F. (2018), Testing Over- and Underidentification in Linear Models, with Applications to Dynamic Panel Data and Asset-Pricing Models, Dept of Economics Discussion Paper 18/696, University of Bristol.