Two-Stage Least Squares as Minimum Distance

Frank Windmeijer

Discussion Paper 17 / 683

7 June 2017
Two-Stage Least Squares as Minimum Distance

Frank Windmeijer*  
Department of Economics and IEU  
University of Bristol, UK  
June 2017

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. This MD estimator is a weighted average of the individual instrument-specific IV estimates, with the weights determined by the variances and covariances of the individual estimators 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

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, page 228). This means that the 2SLS estimator combines the information from the multiple instruments asymptotically optimally under

---

* f.windmeijer@bristol.ac.uk. This research was partly funded by the Medical Research Council (MC_UU_12013/9). I would like to thank Steve Pischke, Mark Schaffer, Chris Skeels and Jon Temple for helpful comments.
these conditions. An alternative estimator is a Minimum Distance (MD) estimator, which is a weighted average of the individual instrument-specific IV estimates of the parameter of interest. It is shown in the next section, Section 2, that the optimal MD estimator with the weights determined by the variances and covariances of the instrument-specific estimators, 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, pages 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

\[ x_i = z'_i \pi + v_i. \]

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 \( P = Z (Z'Z)^{-1} Z' \) and

\[ S(\beta) = (y- x \beta)' P Z (y- x \beta). \]  \(  \) (1)

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

\[ \hat{\beta}_{2SLS} = \arg \min_{\beta} S(\beta) \]

and is given by

\[ \hat{\beta}_{2SLS} = (x' P Z x)^{-1} x' P Z y \]
\[ = (\hat{x}' \hat{x})^{-1} \hat{x}' y, \]  \(  \) (2)

where \( \hat{x} = P Z x. \)

Next consider the individual instrument-specific IV estimates for \( \beta, \)

\[ \hat{\beta}_j = (\hat{x}_j' \hat{x}_j)^{-1} \hat{x}_j' y; \]
\[ \hat{x}_j = P_j x \]

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

\[ \hat{\beta}_{ind} = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_{k_z})'. \]  \(  \) (3)
Under standard regularity assumptions and conditional homoskedasticity, \( E(u^2_i | z_i) = \sigma_u^2 \), the limiting distribution of \( \hat{\beta}_{ind} \) is given by

\[
\sqrt{n} \left( \hat{\beta}_{ind} - \iota \beta \right) \xrightarrow{d} N \left( 0, \sigma_u^2 \Omega \right),
\]

where \( \iota \) is a \( k_z \)-vector of ones. The elements of \( \Omega \) are given by

\[
\Omega_{js} = (\text{plim} (\hat{x}'_j \hat{x}_j / n))^{-1} (\text{plim} (\hat{x}'_j \hat{x}_s / n)) (\text{plim} (\hat{x}'_s \hat{x}_s / n))^{-1}
\]

for \( j, s = 1, \ldots, k_z \). The asymptotic variances of \( \hat{\beta}_j \) and the covariances of \( \hat{\beta}_j \) and \( \hat{\beta}_s \) are therefore given by

\[
\text{Var} \left( \hat{\beta}_j \right) = \sigma_u^2 (\hat{x}'_j \hat{x}_j)^{-1}
\]

\[
\text{Cov} \left( \hat{\beta}_j, \hat{\beta}_s \right) = \sigma_u^2 (\hat{x}'_j \hat{x}_j)^{-1} (\hat{x}'_j \hat{x}_s) (\hat{x}'_s \hat{x}_s)^{-1},
\]

for \( j, s = 1, \ldots, k_z \), \( j \neq s \), resulting in

\[
\text{Var} \left( \hat{\beta}_{ind} \right) = \sigma_u^2 \hat{\Omega};
\]

\[
\hat{\Omega} = \hat{D}^{-1} \hat{X}_{ind}' \hat{X}_{ind} \hat{D}^{-1}, \quad (4)
\]

where

\[
\hat{X}_{ind} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \ldots & \hat{x}_{k_z} \end{bmatrix}; \quad (5)
\]

\[
\hat{D} = \text{diag} (\hat{x}'_j \hat{x}_j), \quad (6)
\]

\( j = 1, \ldots, k_z \), where \( \text{diag} (a_j) \) is a diagonal matrix with \( j \)-th diagonal element \( a_j \). Under standard conditions, \( \text{plim} \left( n \hat{\Omega} \right) = \Omega \).

Let

\[
Q(\beta) = \left( \hat{\beta}_{ind} - \iota \beta \right)' \hat{\Omega}^{-1} \left( \hat{\beta}_{ind} - \iota \beta \right). \quad (7)
\]

The optimal MD estimator for \( \beta \) is then defined as

\[
\hat{\beta}_{md} = \arg \min_{\beta} Q(\beta),
\]

resulting in

\[
\hat{\beta}_{md} = \left( \iota \hat{\Omega}^{-1} \iota \right)^{-1} \iota \hat{\Omega}^{-1} \hat{\beta}_{ind}. \quad (8)
\]
It is clear that the MD estimator is a weighted average of the individual instrument specific estimates,

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

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

The 2SLS estimator is also a weighted average. Let

$$\hat{\Delta} = \text{diag}\left(\left(z_j' z_j\right)^{-1} z_j' x\right),$$

$j = 1,\ldots, k_z$, then

$$\hat{X}_{ind} = Z \hat{\Delta}.$$

As

$$\hat{\beta}_{ind} = D^{-1} \hat{X}'_{ind} y,$$

it therefore follows that

$$\hat{\beta}_{2sls} = \left(x' P_Z x\right)^{-1} x' P_Z y = \left(x' P_Z x\right)^{-1} x' Z (Z' Z)^{-1} \hat{\Delta}^{-1} D^{-1} \hat{\Delta} Z' y = \left(x' P_Z x\right)^{-1} x' Z (Z' Z)^{-1} \hat{\Delta}^{-1} D \hat{\beta}_{ind} = \sum_{j=1}^{k_z} w_{2sls,j} \hat{\beta}_j.$$

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

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

**Proof.** As $\hat{x}' \hat{x}_j = x' P_Z z_j x = x' P_Z x = x' \hat{x}_j = \hat{x}' \hat{x}_j$, it follows that

$$t' \hat{D} = \begin{pmatrix} \hat{x}'_1 \hat{x}_1 & \hat{x}'_2 \hat{x}_2 & \ldots & \hat{x}'_{k_z} \hat{x}_{k_z} \end{pmatrix} = \hat{x}' \hat{X}_{ind} = x' \hat{X}_{ind},$$

and hence

$$\hat{D}^{-1} \hat{X}'_{ind} \hat{x} = \hat{D}^{-1} \hat{X}'_{ind} x = t.$$
As
\[ P_{\hat{X}_{ind}} = Z\hat{\Delta} (\Delta Z'Z\hat{\Delta})^{-1} Z\hat{\Delta} = P_Z, \]
it follows that, for \( \beta \in \mathbb{R} \),
\[
S (\beta) = (y - x\beta)' P_Z (y - x\beta) \]
\[
= (y - x\beta)' P_{\hat{X}_{ind}} (y - x\beta) \]
\[
= (y - x\beta)' \hat{X}_{ind} \hat{D}^{-1} \hat{D} (\hat{X}_{ind}'\hat{X}_{ind})^{-1} \hat{D} \hat{D}^{-1} \hat{X}_{ind}' (y - x\beta) \]
\[
= (\hat{D}^{-1} \hat{X}_{ind}' y - \hat{D}^{-1} \hat{X}_{ind}' x\beta)' \hat{\Omega}^{-1} (\hat{D}^{-1} \hat{X}_{ind}' y - \hat{D}^{-1} \hat{X}_{ind}' x\beta) \]
\[
= (\hat{\beta}_{md} - \epsilon\beta)' \hat{\Omega}^{-1} (\hat{\beta}_{md} - \epsilon\beta) \]
\[
= Q (\beta). \]

Hence \( \hat{\beta}_{md} = \hat{\beta}_{2SLS} \). \( \blacksquare \)

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
\[
Sar (\hat{\beta}_{2sls}) = \hat{\sigma}_u^{-2} S (\hat{\beta}_{2sls}) ,
\]
where \( \hat{\sigma}_u^2 = (y - x\hat{\beta}_{2sls})' (y - x\hat{\beta}_{2sls}) / n \). Under the null, standard regularity assumptions and conditional homoskedasticity, \( Sar (\hat{\beta}_{2sls}) \) converges in distribution to a \( \chi^2_{k_z-1} \) distributed random variable, see e.g. Hayashi (2000, page 228).

Next consider the MD criterion
\[
MD (\hat{\beta}_{md}) = \hat{\sigma}_u^{-2} Q (\hat{\beta}_{md}) ,
\]
where we can use \( \hat{\sigma}_u^2 \) because \( \hat{\beta}_{md} = \hat{\beta}_{2sls} \). Denote \( \beta_j = \plim (\hat{\beta}_j) \) for \( j = 1, ..., k_z \). Then under the null hypothesis \( H_0 : \beta_1 = \beta_2 = ... = \beta_{k_z} = \beta \), and the same maintained assumptions as above, \( MD (\hat{\beta}_{md}) \) converges in distribution to a \( \chi^2_{k_z-1} \) distributed random variable, see e.g. Cameron and Trivedi (2005, page 203).

It follows directly from the results of Proposition 1 that \( S (\hat{\beta}_{2sls}) = Q (\hat{\beta}_{md}) \) and hence
\[
Sar (\hat{\beta}_{2sls}) = MD (\hat{\beta}_{md}) .
\]
Remark 1 The equivalence of $\hat{\beta}_{2sls}$ and $MD \hat{\beta}_{md}$ establishes an intuitive interpretation of the Sargan test. It tests whether the individual instrument-specific estimators all estimate the same parameter value, see also Parente and Santos Silva (2012).

2.2 Efficient Two-Step Estimation

The results above were derived assuming conditional homoskedasticity. The equivalence results can be extended to the efficient two-step GMM estimator. For the cross-sectional setup considered here, this would cover the case of general conditional heteroskedasticity, $E(u_i^2|z_i) = g(z_i)$. Using $\hat{\beta}_{2sls}$ as the initial consistent one-step GMM estimator, the efficient two-step GMM estimator is defined as

$$\hat{\beta}_{gmm} = \arg \min_{\beta} J(\beta),$$  \hspace{1cm} (9)

where

$$J(\beta) = (y - x\beta)' Z W^{-1} \left( \hat{\beta}_{2sls} \right) Z' (y - x\beta);$$ \hspace{1cm} (10)

$$W(\hat{\beta}_{2sls}) = \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_{2sls} \right)^2 z_i' z_i.$$ \hspace{1cm} (11)

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

Next, consider the MD criterion based on a robust variance estimator for $\hat{\beta}_{ind}$, using $\hat{\beta}_{md}$, defined as

$$\hat{\Omega}_r = \hat{D}^{-1} \hat{\Sigma} \hat{D}^{-1},$$ \hspace{1cm} (12)

where the elements of $\hat{\Sigma}$ are given by

$$\hat{\Sigma}_{j,s} = \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_{md} \right)^2 \hat{x}_{ji} \hat{x}_{si},$$ \hspace{1cm} (13)

for $j, s = 1, ..., k_z$.

Define the robust MD estimator as

$$\hat{\beta}_{md,r} = \arg \min_{\beta} MD_r(\beta),$$ \hspace{1cm} (14)

where

$$MD_r(\beta) = \left( \hat{\beta}_{ind} - i \beta \right)' \hat{\Omega}_r^{-1} \left( \hat{\beta}_{ind} - i \beta \right).$$ \hspace{1cm} (15)
Under the null as specified above, \( H_0 : \beta_1 = \beta_2 = \ldots = \beta_{k_z} = \beta \), and standard assumptions, \( MD_r(\hat{\beta}_{md,r}) \) converges in distribution to a \( \chi^2_{k_z-1} \) distributed random variable.

The next proposition establishes the equivalence of \( J(\beta) \) and \( MD_r(\beta) \), hence \( \hat{\beta}_{gmm} = \hat{\beta}_{md,r} \) and \( J(\hat{\beta}_{gmm}) = MD_r(\hat{\beta}_{md,r}) \).

**Proposition 2** Let \( J(\beta), MD_r(\beta), \hat{\beta}_{gmm} \) and \( \hat{\beta}_{md,r} \) be as defined in (10), (15), (9) and (14) respectively. Then, for \( \beta \in \mathbb{R} \), \( J(\beta) = MD_r(\beta) \). Hence \( \hat{\beta}_{md,r} = \hat{\beta}_{gmm} \) and \( MD_r(\hat{\beta}_{md,r}) = J(\hat{\beta}_{gmm}) \).

**Proof.** Denoting the \( i \)-th row of \( \hat{X}_{ind} \) by \( \hat{X}_{ind[i]} \), then \( \hat{\Omega}_r \) can be written as

\[
\hat{\Omega}_r = \hat{D}^{-1} \left( \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_{md} \right)^2 \hat{X}_{ind[i]}' \hat{X}_{ind[i]} \right) \hat{D}^{-1}.
\]

For \( \beta \in \mathbb{R} \),

\[
J(\beta) = (y - x\beta)' Z W^{-1} \left( \hat{\beta}_{2sls} \right) Z' (y - x\beta)
= (y - x\beta)' Z \hat{\Delta} \hat{D}^{-1} \hat{D} \hat{\Delta}^{-1} W^{-1} \left( \hat{\beta}_{2sls} \right) \hat{\Delta}^{-1} \hat{D} \hat{D}^{-1} \hat{\Delta} Z' (y - x\beta)
= (y - x\beta)' \hat{X}_{ind} \hat{D}^{-1} \hat{D} \left( \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_{2sls} \right)^2 \hat{\Delta} \hat{z}_i \hat{\Delta} \right)^{-1} \hat{D} \hat{D}^{-1} \hat{X}_{ind}' (y - x\beta)
= (\hat{\beta}_{ind} - t\beta)' \hat{D} \left( \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_{md} \right)^2 \hat{X}_{ind[i]}' \hat{X}_{ind[i]} \right)^{-1} \hat{D} \left( \hat{\beta}_{ind} - t\beta \right)
= (\hat{\beta}_{ind} - t\beta)' \hat{\Omega}_r^{-1} (\hat{\beta}_{ind} - t\beta)
= MD_r(\beta).
\]

Hence \( \hat{\beta}_{gmm} = \hat{\beta}_{md,r} \) and \( J(\hat{\beta}_{gmm}) = MD_r(\hat{\beta}_{md,r}) \). ■

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

\[
\hat{\Omega}_\Phi = \hat{D}^{-1} \hat{\Phi} \hat{D}^{-1},
\]

with the elements of \( \hat{\Phi} \) given by

\[
\hat{\Phi}_{j,s} = \sum_{i=1}^{n} \left( y_i - x_i \hat{\beta}_j \right) \left( y_i - x_i \hat{\beta}_s \right) \hat{x}_{ji} \hat{x}_{si},
\]

for \( j, s = 1, \ldots, k_z \). The resulting minimum distance estimator, \( \hat{\beta}_{md,\Phi} \), has the same limiting distribution as \( \hat{\beta}_{md,r} \), but differs in finite samples.
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

\[
\hat{\beta}_{2sls} = \text{arg min}_\beta S(\beta)
\]

\[
S(\beta) = (y - X\beta)' P_Z (y - X\beta)
\]

(16)

and is given by

\[
\hat{\beta}_{2sls} = (X' P_Z X)^{-1} X' P_Z y
\]

\[
= (\hat{X}' \hat{X})^{-1} \hat{X}' y
\]

where \( \hat{X} = P_Z X \).

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 \( \hat{\Omega} \). 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 \( \beta_l, l = 1, \ldots, k_x \). Denote by \( \hat{X}_l \) the \( l \)-th column of \( \hat{X} \), and let \( \hat{X}_{-l} \) be the \( k_x - 1 \) columns of \( \hat{X} \), excluding \( \hat{X}_l \). The 2SLS estimator for \( \beta_l \) is given by

\[
\hat{\beta}_{l,2sls} = (\hat{X}_l' M_{\hat{X}_{-l}} \hat{X}_l)^{-1} \hat{X}_l' M_{\hat{X}_{-l}} y
\]

\[
= (\bar{x}_l' \bar{x}_l)^{-1} \bar{x}_l y
\]

(17)

where for a general \( n \times k \) matrix \( A \), \( M_A = I_n - A (A' A)^{-1} A' \), with \( I_n \) is the identity matrix of order \( n \), and

\[
\bar{x}_l = M_{\bar{X}_{-l}} \hat{X}_l.
\]

(18)

Let \( \{ Z^{[t]} \}_{t=1}^{k_z-k_x+1} \) 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, \( \{ Z^{[t]} = (z_t, \ldots, z_{t+k_x-1}) \}_{t=1}^{k_z-k_x+1} \) is such
a set. From these sets, we get \( k_z - k_x + 1 \) just identified IV estimates \( \widehat{\beta}_t^{[l]} \) of \( \beta \). Let 

\[
\widehat{\beta}_{t,\text{ind}} = \left( \widehat{\beta}_t^{[l]} \right)
\]

be the \((k_z - k_x + 1)\)-vector of the individual estimates of \( \beta_t \). Let 

\[
\widehat{X}_t^{[l]} = P_{Z_t} X_t,
\]

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

\[
\widehat{\beta}_{t,\text{ind}}^{[l]} = \left( \widehat{X}_t^{[l]}' M_{\widehat{X}_t^{[l]}} \widehat{X}_t^{[l]} \right)^{-1} \widehat{X}_t^{[l]}' M_{\widehat{X}_t^{[l]}} y
\]

for \( t = 1, \ldots, k_z - k_x + 1 \), where 

\[
\bar{x}_t^{[l]} = M_{\bar{x}_t^{[l]}} \widehat{X}_t^{[l]}.
\] (19)

The asymptotic variances and covariances under conditional homoskedasticity are given by 

\[
\text{Var} \left( \widehat{\beta}_{t,\text{ind}}^{[l]} \right) = \sigma_u^2 \left( \bar{x}_t^{[l]}' \bar{x}_t^{[l]} \right)^{-1}
\]

\[
\text{Cov} \left( \widehat{\beta}_{t,\text{ind}}^{[l]}, \widehat{\beta}_{t,\text{ind}}^{[s]} \right) = \sigma_u^2 \left( \bar{x}_t^{[l]}' \bar{x}_t^{[l]} \right)^{-1} \left( \bar{x}_t^{[l]}' \bar{x}_t^{[l]} \right)^{-1} \left( \bar{x}_t^{[s]}' \bar{x}_t^{[s]} \right)^{-1}.
\]

The MD estimator for \( \beta_t \) is then obtained as 

\[
\widehat{\beta}_{t,\text{ind}} = \arg \min_{\beta_t} Q (\beta_t);
\]

\[
Q (\beta_t) = (\widehat{\beta}_{t,\text{ind}} - \iota \beta_t)' \hat{\Omega}_t^{-1} (\widehat{\beta}_{t,\text{ind}} - \iota \beta_t)
\] (20)

where \( \iota \) is here a \( k_z - k_x + 1 \) vector of ones, and 

\[
\hat{\Omega}_t = \tilde{D}_t^{-1} \bar{x}_t' \bar{x}_t \tilde{D}_t^{-1},
\]

with 

\[
\bar{x}_t = \left( \bar{x}_t^{[1]}, \ldots, \bar{x}_t^{[k_z - k_x + 1]} \right)
\] (21)

and \( \tilde{D}_t = \text{diag} \left( \bar{x}_t^{[1]}' \bar{x}_t^{[1]} \right), \ t = 1, \ldots, k_z - k_x + 1. \)

\( \widehat{\beta}_{t,\text{ind}} \) is therefore given by 

\[
\widehat{\beta}_{t,\text{ind}} = \left( \iota \tilde{D}_t \left( \bar{x}_t' \bar{x}_t \right)^{-1} \tilde{D}_t \right)^{-1} \left( \iota \tilde{D}_t \left( \bar{x}_t' \bar{x}_t \right)^{-1} \tilde{D}_t \right) \widehat{\beta}_{t,\text{ind}}
\]

\[
= \left( \iota \tilde{D}_t \left( \bar{x}_t' \bar{x}_t \right)^{-1} \tilde{D}_t \right)^{-1} \iota \tilde{D}_t \left( \bar{x}_t' \bar{x}_t \right)^{-1} \bar{x}_t y,
\] (22)
where the latter equality holds as
\[ \hat{\beta}_{l,ind} = \tilde{D}_l^{-1} \tilde{X}_l'y. \] (23)

We are now in the same situation as that of the single endogenous variable case, with \( \hat{x}, \) \( \tilde{X}_{ind} \) and \( \tilde{D} \) replaced by \( \tilde{x}_l, \tilde{X}_l \) and \( \tilde{D}_l \) respectively. The next proposition establishes the equivalence of \( \hat{\beta}_{l,2sls} \) and \( \hat{\beta}_{l,md} \) for \( l = 1, \ldots, k_x \).

**Proposition 3** For \( l = 1, \ldots, k_x \), let \( \hat{\beta}_{l,2sls}, \hat{\beta}_{l,ind} \) and \( \hat{\beta}_{l,md} \) be as defined in (17), (23) and (22) respectively, with \( \hat{\beta}_{l,ind} \) based on a collection of \( k_z \) sets of \( k_x \) instruments \( \{Z_t^k\}_{t=1}^{k_z-k_x+1} \) that contains all instruments. Then \( \hat{\beta}_{l,2sls} = \hat{\beta}_{l,md} \) for \( l = 1, \ldots, k_x \).

**Proof.** From the definitions of \( \tilde{x}_l \) and \( \tilde{x}_l^{[t]} \) in (18) and (19) respectively, it follows that
\[
\begin{align*}
\tilde{x}_l^{[t]} &= x_l' \left( P_Z - P_Z X_{-l} \left( X_{-l}' P_Z X_{-l} \right)^{-1} X_{-l}' P_Z \right) \\
&\quad \left( P_Z[t] - P_Z[t] X_{-l} \left( X_{-l}' P_Z[t] X_{-l} \right)^{-1} X_{-l}' P_Z[t] \right) x_l \\
&= x_l' \left( P_Z[t] - P_Z[t] X_{-l} \left( X_{-l}' P_Z[t] X_{-l} \right)^{-1} X_{-l}' P_Z[t] \right) x_l \\
&= x_l' \tilde{x}_l^{[t]} = \tilde{x}_l^{[t]} \tilde{x}_l^{[t]},
\end{align*}
\]
for \( t = 1, \ldots, k_z-k_x+1 \), and hence
\[ i'D_l = \tilde{x}_l \tilde{x}_l. \]

Therefore, from (22),
\[ \hat{\beta}_{l,md} = \left( \tilde{x}_l' P_{\tilde{X}_l} \tilde{x}_l \right)^{-1} \tilde{x}_l' P_{\tilde{X}_l} y. \]

As the sets of instruments \( \{Z_t^k\}_{t=1}^{k_z-k_x+1} \) contain all \( k_z \) instruments, it follows that \( \tilde{x}_l \) is in the column space of \( \tilde{X}_l \), and so \( P_{\tilde{X}_l} \tilde{x}_l = \tilde{x}_l \). Therefore,
\[ \hat{\beta}_{l,md} = (\tilde{x}_l \tilde{x}_l)^{-1} \tilde{x}_l y = \hat{\beta}_{l,2sls}, \]
for \( l = 1, \ldots, k_x \). \( \blacksquare \)

Next, consider the Sargan test statistic, given by
\[ Sar \left( \hat{\beta}_{2sls} \right) = \hat{\sigma}_u^{-2} \left( y - X\hat{\beta}_{2sls} \right)' P_Z \left( y - X\hat{\beta}_{2sls} \right), \] (24)
where \( \hat{\sigma}_u^2 = \left( y - X\hat{\beta}_{2sls} \right)' \left( y - X\hat{\beta}_{2sls} \right) / n \). Under the null \( H_0 : E(z_iu_i) = 0 \), standard regularity conditions and conditional homoskedasticity, \( \text{Sar} \left( \hat{\beta}_{2sls} \right) \xrightarrow{d} \chi^2_{k_x - k_x} \).

Consider the MD statistics

\[
MD \left( \hat{\beta}_{l,md} \right) = \hat{\sigma}_u^{-2} \left( \hat{\beta}_{l,ind} - t \beta_l \right)' \hat{\Omega}_l^{-1} \left( \hat{\beta}_{l,ind} - t \beta_l \right),
\]

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

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

**Proposition 4** Let \( \text{Sar} \left( \hat{\beta}_{2sls} \right) \) and \( MD \left( \hat{\beta}_{l,md} \right) \) be defined as in (24) and (25), then \( \text{Sar} \left( \hat{\beta}_{2sls} \right) = MD \left( \hat{\beta}_{l,md} \right) \) for \( l = 1, \ldots, k_x \).

**Proof.** As \( \hat{X}' \left( y - X\hat{\beta}_{2sls} \right) = 0, \hat{\beta}_{l,2sls} = \hat{\beta}_{l,md}, \) and defining \( \bar{y} = M_{\bar{X}_{-l}}Pzy, \) it follows that

\[
S \left( \hat{\beta}_{2sls} \right) = \left( y - X\hat{\beta}_{2sls} \right)' P_Z \left( y - X\hat{\beta}_{2sls} \right)
= \left( y - x_i\hat{\beta}_{l,2sls} - X_{-l}\hat{\beta}_{l,2sls} \right)' P_Z \left( y - x_i\hat{\beta}_{l,2sls} - X_{-l}\hat{\beta}_{l,2sls} \right)
= \left( y - x_i\hat{\beta}_{l,2sls} - X_{-l}\hat{\beta}_{l,2sls} \right)' P_Z M_{\bar{X}_{-l}} P_Z \left( y - x_i\hat{\beta}_{l,2sls} - X_{-l}\hat{\beta}_{l,2sls} \right)
= \left( \bar{y} - x_i\hat{\beta}_{l,2sls} \right)' P_{\bar{X}_l} \left( \bar{y} - x_i\hat{\beta}_{l,2sls} \right)
= \left( \bar{y} - x_i\hat{\beta}_{l,2sls} \right)' P_{\bar{X}_l} \left( \bar{y} - x_i\hat{\beta}_{l,2sls} \right)
= \left( \bar{X}_l' \left( \bar{y} - \hat{X}_l' \hat{\beta}_{l,2sls} \right) \right)' \hat{D}_l \left( \bar{X}_l' \hat{X}_l \right)^{-1} \hat{D}_l \left( \bar{D}_l' \bar{X}_l' \bar{y} - \bar{D}_l' \bar{X}_l' \hat{\beta}_{l,md} \right)
= \left( \hat{\beta}_{l,ind} - t \hat{\beta}_{l,md} \right)' \hat{\Omega}_l^{-1} \left( \hat{\beta}_{l,ind} - t \hat{\beta}_{l,md} \right)
= Q \left( \hat{\beta}_{l,md} \right)
\]

and hence \( \text{Sar} \left( \hat{\beta}_{2sls} \right) = MD \left( \hat{\beta}_{l,md} \right) \) for \( l = 1, \ldots, k_x \). \( \blacksquare \)

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


