Università di Pavia

Macroeconometria

Instrumental variables
and
the Generalized Method of Moments

Eduardo Rossi
Università di Pavia
Introduction There is a problem of endogeneity in the linear model

\[ y_t = x'_t \beta + \varepsilon_t \]

if \( \beta \) is the parameter of interest and \( E (x_t \varepsilon_t) \neq 0 \).

Example Measurement error in the regressor.

Suppose that \((y_t, x^*_t)\) are joint random variables

\[ E (y_t | x_t) = x^*_t \beta \]

is linear. But \( x^*_t \) is not observed. Instead we observe

\[ x_t = x^*_t + u_t \]
\[ u_t \ (k \times 1) \]

measurement error, independent of \( y_t \) and \( x_t^* \):

\[
y_t = x_t^* \beta + \varepsilon_t \\
= (x_t - u_t) \beta + \varepsilon_t \\
= x_t' \beta + v_t
\]

where

\[ v_t = \varepsilon_t - u_t' \beta. \]

The problem is that

\[
E (x_t v_t) = E \left[ (x_t^* + u_t) (\varepsilon_t - u_t' \beta) \right] = -E \left( u_t u_t' \right) \beta \neq 0
\]
If $\beta \neq 0$ and $E(u_t u'_t) \neq 0$. It follows that if $\hat{\beta}$ is the OLS estimator

$$\hat{\beta} = \beta - \left[ \sum_t (x_t x'_t) \right]^{-1} \sum_t x_t v_t$$

then

$$p \lim \hat{\beta} = \beta - p \lim \left\{ \left[ \frac{1}{N} \sum_t (x_t x'_t) \right]^{-1} \frac{1}{N} \sum_t x_t v_t \right\}$$

$$p \lim \hat{\beta} = \beta - (E(x_t x'_t))^{-1} E(u_t u'_t) \beta \neq \beta.$$
**Instrumental Variables Estimator** Let the equation of interest be

\[ y_t = x_t'\beta + \varepsilon_t \]  \hspace{1cm} (1)

where \( x_t \) \((k \times 1)\), \( E(x_t\varepsilon_t) \neq 0 \) so that there is a problem of endogeneity. Eq. (1) is called *structural equation*. In matrix notation, this can be written as

\[ y = X\beta + \varepsilon \]

any solution to the problem of endogeneity requires additional information which we call *instruments*. 
**Definition** The vector $z_t$, $(l \times 1)$, is an instrumental variable for eq.(1) if $E(z_t \varepsilon_t) = 0$.

In a typical set-up, some regressors in $x_t$ will be uncorrelated with $\varepsilon_t$ (for example, at least the intercept). Thus we make the partition

$$x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} k_1$$

where $E(x_{1t} \varepsilon_t) = 0$ yet $E(x_{2t} \varepsilon_t) \neq 0$. We call $x_{1t}$ exogenous and $x_{2t}$ endogenous. $x_{1t}$ is an instrumental for eq.(1).
So we have the partition

\[ z_t = \begin{bmatrix} x_{1t} \\ z_{2t} \end{bmatrix} k_1 \]

Three possible cases:

Just-identified if \( l = k \)

Over-identified if \( l > k \)

Not identified if \( l < k \).
Reduced Form The reduced form relationship between the variables or "regressors" $x_t$ and the instruments $z_t$ is found by linear projection. Let

$$\Gamma = E\left(z_t z_t'\right)^{-1} E(z_t x_t')$$

be the $(l \times k)$ matrix of coefficients from a projection of $x_t$ on $z_t$, and define

$$u_t' = x_t' - z_t' \Gamma$$

as the projection error. Then the reduced form linear relationship between $x_t$ and $z_t$ is

$$x_t = \Gamma' z_t + u_t$$

(2)
In matrix notation

\[ X = Z\Gamma + u \]  \hspace{1cm} (3)

\( u (n \times k) \). By construction

\[ E(z_tu'_t) = 0 \]

So (3) is a projection and can be estimated by OLS

\[ X = Z\hat{\Gamma} + \hat{u} \]
\[ \hat{\Gamma} = (Z'Z)^{-1}Z'X \]

Substituting \( X = Z\Gamma + u \) in \( y = X\beta + \varepsilon \)

\[ y = (Z\Gamma + u)\beta + \varepsilon \]
\[ = Z\lambda + v \]  \hspace{1cm} (4)

where \( \lambda = \Gamma\beta, \ v = u\beta + \varepsilon \).
Observe that

\[ E(z_t v_t) = E(z_t u_t') \beta + E(z_t \varepsilon_t) = 0 \]

Thus (4) is a projection equation and may be estimated by OLS. This is

\[ y = Z\hat{\lambda} + \hat{v} \]

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

The equation (4) is the reduced form for \( y \). The system

\[ y = Z\lambda + v \]
\[ X = Z\Gamma + u \]

OLS yields the reduced-form estimates \((\hat{\lambda}, \hat{\Gamma})\).
Identification The structural parameter $\beta$ relates to $(\lambda, \Gamma)$ through $\lambda = \Gamma \beta$. The parameter is identified, meaning that it can be recovered from the reduced form, if

$$\text{rank} (\Gamma) = k \quad (5)$$

Assume that (5) holds.

If $l = k$, then $\beta = \Gamma^{-1} \lambda$. If $l > k$, then for any $W > 0$, $\beta = (\Gamma' W \Gamma)^{-1} \Gamma' W \lambda$. If (5) is not satisfied, then $\beta$ cannot be recovered from $(\lambda, \Gamma)$.

Note that a necessary (although not sufficient) condition for (5) is $l \geq k$. 
Since $X$ and $Z$ have the common variables $X_1$, we can rewrite some of the expressions

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$$
$$Z = \begin{bmatrix} X_1 & Z_2 \end{bmatrix}$$

we can partition $\Gamma$ as

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} I_{k_1} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix} \begin{bmatrix} l_1 \times k_2 \\ l_2 \times k_2 \end{bmatrix}$$
\[ Z\Gamma = \begin{bmatrix} X_1 & Z_2 \end{bmatrix} \begin{bmatrix} I & \Gamma_{12} \\ 0 & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} X_1 & X_1\Gamma_{12} + Z_2\Gamma_{22} \end{bmatrix} \]

\( \beta \) is identified if \( \text{rank}(\Gamma) = k \), which is true if and only if \( \text{rank}(\Gamma_{22}) = k_2 \) (by the upper-diagonal structure of \( \Gamma \)). The key to identification of the model rests on the \((l_2 \times k_2)\) matrix \( \Gamma_{22} \).
Instrumental Variable Estimation

Suppose that the model is just identified \((l = k)\). Then \(\beta = \Gamma^{-1} \lambda\). This suggests the *Indirect Least Squares* estimator:

\[
\hat{\beta}_{IV} = \hat{\Gamma}^{-1} \hat{\lambda}
\]

\[
= \left[ (Z'Z)^{-1} Z'X \right]^{-1} \left[ (Z'Z)^{-1} Z'y \right]
\]

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

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

\(\hat{\beta}_{IV}\) is the instrumental variables estimator of \(\beta\).
Since \((\hat{\lambda}, \hat{\Gamma}) \xrightarrow{p} (\lambda, \Gamma)\) and \(\Gamma\) is invertible, \(\beta\) is consistent. A more direct way to see consistency is to substitute the equation \(y = X\beta + \varepsilon\) into the equation for \(\hat{\beta}\) to obtain

\[
\hat{\beta}_{IV} = (Z'X)^{-1} (Z' (X\beta + \varepsilon))
= \beta + (Z'X)^{-1} Z'\varepsilon
\]

\[(Z'X)^{-1} Z'\varepsilon \xrightarrow{p} 0\]

if \(E(z_tx'_t)\) is invertible and \(E(x_t\varepsilon_t) = 0\). These are the identifying assumptions.
The asymptotic distribution of IV estimator

\[ Q_{ZX} = p \lim \left( \frac{Z'X}{n} \right) \]

\[ Q_{ZZ} = p \lim \left( \frac{Z'Z}{n} \right) \]

\[ \sqrt{n} (\hat{\beta}_{IV} - \beta) = \sqrt{n} \left[ (Z'X)^{-1} Z'\varepsilon \right] \]

Under the hypothesis that \( E(\varepsilon \varepsilon' | X, Z) = \sigma^2 I_n \), The Central Limit Theorem gives us the limit distribution

\[ \sqrt{n} (\hat{\beta}_{IV} - \beta) \xrightarrow{L} N \left( 0, \sigma^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1} \right) \]

The estimate of \( \sigma^2 \)

\[ \hat{\sigma}^2 = \frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) \]
The Method of Moments

The assumption

\[ E(\epsilon_t | x_t) = 0 \]

is more restrictive than

\[ E(\epsilon_tx_t) = 0. \]

In fact given

\[ E(\epsilon_t | x_t) = 0 \rightarrow E(\epsilon_tx_t) = 0. \]

The Linear Regression Model satisfies \( E(x_t\epsilon_t) = 0 \). Let \( \beta_0 \) denote the true value of \( \beta \), where the latter denotes a generic value of the parameter. Then

\[ E \left[ x_t \left( y_t - x_t'\beta_0 \right) \right] = 0 \]
Another way is to define the "moment function"

\[ g_t(\beta) = x_t(y_t - x'_t\beta) \]

and observe that

\[ E(g_t(\beta_0)) = 0. \]

For values \( \beta \neq \beta_0 \), \( E(g_t(\beta)) \neq 0 \). The empirical, or sample, analog of a moment \( E(x) \) is the sample moment

\[ \frac{1}{n} \sum_{i=1}^{n} x_i. \]
Similarly the empirical analogue of \( E(g_t(\beta)) \) is

\[
\bar{g}_n(\beta) = \frac{1}{n} \sum_{t=1}^{n} g_t(\beta) \\
= \frac{1}{n} \sum_{t=1}^{n} x_t (y_t - x_t' \beta) \\
= \frac{1}{n} (X'y - X'X\beta)
\]

which implies

\[
X'y - X'X\hat{\beta} = 0
\]

these are the "normal equations".
The solution is the **MME** (Method of Moments Estimator) of $\beta$. Note that by definition

$$g_n(\hat{\beta}) = \frac{1}{n}X'\hat{\varepsilon}$$

thus $X'\hat{\varepsilon} = 0$. 

The IV Estimator as a MM Estimator The defining characteristic of $z_t$ as an instrumental variable is that $E(z_t \epsilon_t) = 0$. This suggests the moment equation

$$g_t(\beta) = z_t \left( y_t - x_t' \beta \right)$$

so that at the true value we have the equality

$$E(g_t(\beta_0)) = 0.$$ 

The sample analog is

$$g_n(\beta) = \frac{1}{n} \sum_{t=1}^{n} g_t(\beta)$$

$$= \frac{1}{n} \sum_{t=1}^{n} z_t \left( y_t - x_t' \beta \right)$$

$$= \frac{1}{n} \left( Z'y - Z'X\beta \right)$$
The MME $\hat{\beta}$ sets $g_n(\hat{\beta}) = 0$

$$\left( Z' y - Z' X \hat{\beta} \right) = 0$$

$$\hat{\beta} = (Z' X)^{-1} (Z' y) .$$
**The GMM Estimator** In the overidentified case ($l > k$), the simple IV estimator does not exist. We generalize the MME,

$$
\bar{g}_n(\beta) = 0 \quad (l \times 1) \hspace{1cm} \beta \quad (k \times 1)
$$

so there is no $\hat{\beta}$ such that $\bar{g}_n(\hat{\beta}) = 0$.

The idea of GMM is to set this vector "close" to zero. For some $W_n$ ($l \times l$), let

$$
J(\beta) = n\bar{g}_n(\beta)' W_n \bar{g}_n(\beta)
$$

This is a non-negative measure of the "length" of the vector $\bar{g}_n(\beta)$. 
For example, if $W_n = I$, then

$$J(\beta) = n\bar{g}_n(\beta)' \bar{g}_n(\beta)$$

$$= n \|\bar{g}_n(\beta)\|^2$$

the square of Euclidean length. The GMM estimator minimizes $J(\beta)$.

$$\hat{\beta}_{GMM} = \arg \min_{\beta} J(\beta).$$

Note that if $l = k$, then $\bar{g}_n(\hat{\beta}_{IV}) = 0$ so $J(\hat{\beta}_{IV}) = 0$ and thus $\hat{\beta}_{GMM} = \hat{\beta}_{IV}$. 
The GMM estimator in the linear model The FOC for the definition of the GMM estimator are:

\[
\frac{\partial}{\partial \beta} J(\hat{\beta}) = 0
\]

\[
\frac{\partial}{\partial \beta} \left[ n\bar{g}_n(\hat{\beta})' W_n\bar{g}_n(\hat{\beta}) \right] = 0
\]

\[
\bar{g}_n(\hat{\beta})' W_n \frac{\partial \bar{g}_n(\hat{\beta})}{\partial \beta'} + \bar{g}_n(\hat{\beta})' W_n \frac{\partial \bar{g}_n(\hat{\beta})}{\partial \beta'} = 0
\]

\[
2 \frac{\partial \bar{g}_n(\hat{\beta})'}{\partial \beta} W_n \bar{g}_n(\hat{\beta}) = 0
\]
but in the case of linear model,

$$\bar{g}_n(\hat{\beta}) = n^{-1} (Z'y - Z'X\hat{\beta}) = 0$$

$$\frac{\partial \bar{g}_n(\hat{\beta})}{\partial \beta}' = -\frac{1}{n}X'Z$$

$$-2\frac{1}{n} (X'Z) W_n (Z'y - Z'X\hat{\beta}) \frac{1}{n} = 0$$

$$\hat{\beta}_{GMM} = \left[ X'Z W_n Z'X \right]^{-1} (X'Z) W_n (Z'y)$$

while the estimator depends on $W_n$, the dependence is only up to scale, for if $W_n$ is multiplied by $c$, $cW_n$, for some $c > 0$, $\hat{\beta}_{GMM}$ does not change.
2SLS Estimator Suppose $W_n = (Z'Z)^{-1}$. Then the GMM estimator equals

$$
\hat{\beta}_{2SLS} = \left[ X'Z(Z'Z)^{-1}Z'X \right]^{-1} \left( X'Z(Z'Z)^{-1}Z'y \right)
$$

originally proposed by Theil (1953) and Basmann (1957). It is the classic estimator for linear equations with instruments. Observe that

$$
P_z = Z(Z'Z)^{-1}Z'
$$

$$
\hat{X} = P_zX = Z(Z'Z)^{-1}Z'X
$$
\[ \hat{\beta} = \left( X' P_z X \right)^{-1} X' P_z y \]
\[ = \left( X' P'_z P_z X \right)^{-1} X' P'_z y \]
\[ = \left( \hat{X}' \hat{X} \right)^{-1} \hat{X}' y \]

The source of the "two stage" name is since it can be computed as follows:

- First regress \( X \) on \( Z \),

\[ \hat{\Gamma} = \left( Z' Z \right)^{-1} Z' X \]
\[ \hat{X} = Z \hat{\Gamma} = P_z X \]
• Second regress $y$ on $\hat{X}$,

$$\hat{\beta} = (\hat{X}'\hat{X})^{-1}\hat{X}'y.$$
Recall that $X = [X_1, X_2]$ and $Z = [X_1, Z_2]$. Then

\[ \hat{X} = [\hat{X}_1, \hat{X}_2] \]
\[ = [X_1, P_z X_2] \]

since $X_1 = Z_1$. In fact,

\[ \hat{X} = Z \hat{\Gamma} = P_z X \]

\[ \hat{X} = [Z_1, Z_2] \begin{bmatrix} I & \hat{\Gamma}_{12} \\ 0 & \hat{\Gamma}_{22} \end{bmatrix} \]
\[ = [Z_1, Z_1 \hat{\Gamma}_{12} + Z_2 \hat{\Gamma}_{22}] . \]
In the second stage, we regress $y$ on $X_1$ and $\hat{X}_2$. So only the endogenous variables $X_2$ are replaced by their fitted values:

$$\hat{X}_2 = Z_1\hat{\Gamma}_{12} + Z_2\hat{\Gamma}_{22}.$$
Asymptotic Distribution of the GMM Estimator

Assume that $W_n \xrightarrow{p} W > 0$. Let

$$Q = E(x_t z_t')$$

$$\Omega = E(z_t z_t' \varepsilon_t^2) = E(g_t g_t')$$

where $g_t = z_t \varepsilon_t$. The GMM estimator

$$\hat{\beta}_{GMM} = \left[ X' Z W_n Z' X \right]^{-1} (X' Z) W_n (Z' y)$$

$$= \beta + \left[ X' Z W_n Z' X \right]^{-1} (X' Z) W_n (Z' \varepsilon)$$
Then

$$\left( \frac{1}{n} X'Z \right) W_n \left( \frac{1}{n} Z'X \right) \xrightarrow{p} Q'WQ$$

$$\left( \frac{1}{n} X'Z \right) W_n \left( \frac{1}{n} Z'\varepsilon \right) \xrightarrow{L} N \left( 0, Q'W\Omega WQ \right)$$

Theorem

$$\sqrt{n} \left( \hat{\beta}_{GMM} - \beta \right) \xrightarrow{L} N \left( 0, V \right)$$

$$V = \left( Q'WQ \right)^{-1} \left( Q'W\Omega WQ \right) \left( Q'WQ \right)^{-1}.$$. 
The Optimal Weighting Matrix

The optimal weighting matrix $W_0$ is one which minimizes $V$. This turns out to be $W_0 = \Omega^{-1}$. This yields the efficient GMM estimator

$$\hat{\beta}_{GMM} = \left[ X'Z\Omega^{-1}Z'X \right]^{-1} (X'Z) \Omega^{-1} (Z'y)$$

For the efficient GMM estimator

$$\sqrt{n} \left( \hat{\beta}_{GMM} - \beta \right) \overset{L}{\rightarrow} N \left( 0, \left( Q'\Omega^{-1}Q \right)^{-1} \right)$$

This estimator is efficient only in the sense that it is the best (asymptotically) in the class of GMM estimators with this set of moment conditions.
$W_0 = \Omega^{-1}$ is not known in practice, but it can be estimated consistently. For any $W_n \xrightarrow{p} W_0$, we still call $\hat{\beta}_{GMM}$ the efficient GMM estimator as it has the same asymptotic distribution. In the case

$$E(\varepsilon_t^2 | z_t) = \sigma^2$$

then

$$W_0 = \sigma^{-2} \left( E(z_tz_t') \right)^{-1} \propto \left( E(z_tz_t') \right)^{-1}$$

2SLS sets $W_n = (Z'Z)^{-1}$. So under homoskedasticity 2SLS is asymptotically efficient. In general, if $l > k$, then 2SLS is asymptotically inefficient.
Estimation of the Efficient Weight Matrix Let

\[ W_n = \left( \frac{1}{n} \sum \hat{g}_t \hat{g}'_t \right)^{-1} \]

\( \hat{g}_t = z_t \hat{\varepsilon}_t, \hat{\varepsilon}_t \) are consistent first-stage residuals, such as the 2SLS residuals. When

\[ W_n \overset{p}{\rightarrow} \Omega^{-1} = W_0 \]

GMM using \( W_n \) is asymptotically efficient.

In a testing context, the moment conditions may be violated, in which case \( E(\hat{g}_t) = 0 \), so \( W_n \) above will contain the sum of a bias and a variance component, which is undesirable. Solution re-center the moment conditions: \( W_n = \frac{1}{n} \sum_t (\hat{g}_t - \bar{g}_n) (\hat{g}_t - \bar{g}_n)' \), \( \bar{g}_n = \frac{1}{n} \sum_t \hat{g}_t \).
Simple way to compute the efficient GMM estimator:

1. Estimate $\hat{\beta}$ by 2SLS, and construct the residuals

$\hat{\varepsilon}_t = y_t - x_t' \hat{\beta}_{2SLS}$

$\hat{g}_t = z_t \hat{\varepsilon}_t$

$\hat{g} = \begin{bmatrix} \hat{g}_1' \\ \vdots \\ \hat{g}_n' \end{bmatrix} \quad (n \times l)$
2. The efficient GMM estimator is either:

$$
\hat{\beta}_{GMM} = \left[ X'Z (\hat{g}'\hat{g})^{-1} Z'X \right]^{-1} \left[ X'Z (\hat{g}'\hat{g})^{-1} Z'y \right]
$$

or

$$
\hat{\beta}_{GMM} = \left[ X'Z (\hat{g}'\hat{g} - n\bar{g}_n\bar{g}'_n)^{-1} Z'X \right]^{-1} \left[ X'Z (\hat{g}'\hat{g} - n\bar{g}_n\bar{g}'_n)^{-1} Z'y \right].
$$
An estimator of $\text{Var}(\hat{\beta}_{GMM})$:

$$\text{Var}(\hat{\beta}_{GMM}) = \left( X'Z \left( \hat{g}'\hat{g} \right)^{-1} Z'X \right)^{-1}$$

or

$$\text{Var}(\hat{\beta}_{GMM}) = \left( X'Z \left( \hat{g}'\hat{g} - n\bar{g}_n\bar{g}'_n \right)^{-1} Z'X \right)^{-1}$$

Asymptotic standard errors are given by the square roots of the diagonal elements of $\text{Var}(\hat{\beta}_{GMM})$. 
**Over-identification Test** If \( l > k \), there are restrictions which are testable. Under the hypothesis of correct specification, and if the weight matrix is asymptotically efficient,

\[
J = J(\hat{\beta}) \xrightarrow{L} \chi^2_{l-k}
\]

\[
J = n\bar{g}'_n W_n \bar{g}_n
\]

\[
= n^2 \bar{g}'_n \left( \hat{g}' \hat{g} - n\bar{g}_n \bar{g}'_n \right)^{-1} \bar{g}_n
\]

the degrees of freedom of the asymptotic distribution are the number of overidentifying restrictions. If \( J > J^*_{0.05} \) (Critical value) we can reject the model. Based on this information alone, it is unclear what is wrong, but it is typically cause for concern.
GMM: The general case In its most general form, GMM applies whenever an economic or statistical model implies the $l \times 1$ moment conditions

$$E [g_t (\beta)] = 0.$$  

Identification requires $l \geq k = \text{dim} (\beta)$. The GMM minimizes

$$J = n \bar{g}_n \bar{g}_n' W_n \bar{g}_n$$

$$\bar{g}_n (\beta) = \frac{1}{n} \sum_t g_t (\beta)$$

$$W_n = \left( \frac{1}{n} \sum_t \hat{g}_t \hat{g}_t' \right)^{-1}$$
or

$$W_n = \left( \frac{1}{n} \sum_t \hat{g}_t \hat{g}'_t - \bar{g}_n \bar{g}'_n \right)^{-1}$$

with $\hat{g}_t = g_t(\tilde{\beta})$ constructed using $\tilde{\beta}$, a preliminary consistent estimator, perhaps obtained by first setting $W_n = I$. 
Under general regularity conditions,

$$\sqrt{n}(\hat{\beta}_{GMM} - \beta) \xrightarrow{L} N\left(0, (G'\Omega^{-1}G)^{-1}\right)$$

where $$\Omega^{-1} = E(g_tg_t')^{-1}$$ and $$G = E\left[\frac{\partial g_t(\beta)}{\partial \beta'}\right]$$.

The variance of $$\hat{\beta}$$ may be estimated by $$\left(\hat{G}'\hat{\Omega}^{-1}\hat{G}\right)^{-1}$$ where

$$\hat{\Omega} = n^{-1} \sum_t \hat{g}_t\hat{g}_t'$$

or

$$\hat{\Omega} = n^{-1} \sum_t \hat{g}_t\hat{g}_t' - \bar{g}_n\bar{g}_n'$$
and

\[ \hat{G} = n^{-1} \sum_t \frac{\partial}{\partial \beta} g_t(\hat{\beta}) \].