Instrumental Variables and Two-Stage Least Squares

Generalised Least Squares

Professor Menelaos Karanasos

December 2011

Generalised Least Squares:

Assume that the postulated model is

$$y = Xb + e, \tag{1}$$

where

$$e \sim N(0, \sigma^2 \Omega),$$

where Ω is a positive definite matrix-this implies that Ω^{-1} is also a positive definite matrix.
Thus it is possible to find a nonsingular matrix P such that

$$\Omega^{-1} = P'P.$$

or

$$\Omega = (P'P)^{-1} = P^{-1}(P')^{-1}.$$

Premultiply the linear model in equation (1) by P, to obtain

$$Py = PXb + Pe \tag{2}$$

Denote Pe by u. Then

$$\begin{aligned} \text{Var}(u) &= E(uu') = E(Pee'P') \\ &= P\underbrace{E(ee')}_{\sigma^2\Omega}P' = \sigma^2 P \underbrace{\Omega}_{P^{-1}(P')^{-1}} P' \\ &= \sigma^2 PP^{-1}(P')^{-1}P' = \sigma^2 I. \end{aligned}$$

Thus the transformed variables in equation (2) satisfy the conditions under which OLS is BLUE.

The coefficient estimated vector from OLS regression of Py on PX is the generalized least squares (GLS) estimator:

$$\begin{aligned}\beta_G &= [(PX)'PX]^{-1}(PX)'Py \\ &= (X'\underbrace{P'P}_{\Omega^{-1}}X)^{-1}X'P'Py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.\end{aligned}$$

From the OLS theory it follows that

$$\begin{aligned}Var(\beta_G) &= \sigma^2[(PX)'PX]^{-1} = \sigma^2(X'\underbrace{P'P}_{\Omega^{-1}}X)^{-1} \\ &= \sigma^2(X'\Omega^{-1}X)^{-1}.\end{aligned}$$

An unbiased estimator of the unknown σ^2 is readily obtained from the application of OLS to the transformed model. It is

$$\begin{aligned} s^2 &= \widehat{u}'\widehat{u}/N - k \\ &= (Py - PX\beta_G)'(Py - PX\beta_G)/N - k \\ &= (y - X\beta_G)'\underbrace{P'P}_{\Omega^{-1}}(y - X\beta_G)/N - k = \\ & \quad (y - X\beta_G)'\Omega^{-1}(y - X\beta_G)/N - k, \end{aligned}$$

where $\beta_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$.

Note that the procedures outlined so far imply knowledge of Ω. In practice Ω is unknown, and it is important to develop feasible generalized least squares (FGLS).

Finally, note that if $\sigma^2\Omega = V$, where V is a positive definite variance-covariance matrix.

Then, it follows directly that

$$\begin{aligned} \beta_G &= (X' \underbrace{\Omega^{-1}}_{\sigma^2 V^{-1}} X)^{-1} X' \underbrace{\Omega^{-1}}_{\sigma^2 V^{-1}} y = \\ &= (X' V^{-1} X)^{-1} X' V^{-1} y, \end{aligned}$$

and

$$\begin{aligned} Var(\beta_G) &= \sigma^2 (X' \underbrace{\Omega^{-1}}_{\sigma^2 V^{-1}} X)^{-1} \\ &= \sigma^2 (X' V^{-1} X)^{-1}. \end{aligned}$$

Feasible GLS procedure:

Let

$$V = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{bmatrix}.$$

Further, we hypothesize that

$$\sigma_i^2 = a_0 + a_1 z_i^{a_2},\ i = 1, \ldots, N,$$

where z_i is a single variable, possibly one of the regressors, thought to determine the heteroscedasticity.

Because the OLS residuals $\widehat{e} = y - X\beta$ are consistent estimates of e one can run the nonlinear regression

$$\widehat{e}_i^2 = a_0 + a_1 z_i^{a_2} + v_i.$$

Estimates of the disturbance variances are then

$$\widehat{\sigma}_i^2 = \alpha_0 + \alpha_1 z_i^{\alpha_2},\ i = 1, \ldots, N.$$

These estimates give the $\widehat{V}$ matrix and a feasible GLS procedure.

Instrumental variables (IV) estimators:

Consider the relation

$$y_i = bx_i + e_i, \tag{3}$$

where for simplicity the constant term has been dropped. Suppose however, that the observed value x_i can be represented as the sum of the true value $\widetilde{x}_i$ and a random measurement error v_i, that is

$$x_i = \widetilde{x}_i + v_i.$$

In this case the apropriate relation may be

$$y_i = b\widetilde{x}_i + e_i \tag{4}$$

If we assume that equation (4) is the maintained specification but that observations are only available on x_i and not on $\widetilde{x}_i$, what happens if we use OLS?

The OLS slope is

$$\begin{aligned} \beta &= \frac{\sum \underbrace{y}_{b\widetilde{x}_i + e_i} x}{\sum x^2} = \frac{\sum x(b\widetilde{x} + e)}{\sum x^2} \\ &= b\frac{\sum x\widetilde{x}}{\sum x^2} + \frac{\sum xe}{\sum x^2} \Rightarrow \\ E(\beta) &= b\frac{\sum x\widetilde{x}}{\sum x^2} \neq b. \end{aligned}$$

Thus OLS is biased. This is an example of specification error.

In this case one should make use of instrumental variables which are also commonly referred to as instruments.

Suppose that it is possible to find a data matrix Z of order $N \times l$ ($l \geq k$), which posseses two vital properties:

1. The variables in Z are correlated with those of X
2. The variables in Z are (in the limit) uncorrelated with the disturbance term e.

Premultiplying the general relation by Z' gives

$$\underbrace{Z'y}_{y^*} = \underbrace{Z'X}_{X^*}b + \underbrace{Z'e}_{e^*},$$

with

$$\begin{aligned} Var(e^*) &= (Z'e) = E[(Z'e)(Z'e)'] \\ &= E(Z'ee'Z) \\ &= Z'E(ee')Z = \sigma^2 Z'Z. \end{aligned}$$

This suggests the use of the GLS. The resultant estimator is

$$\begin{aligned}\beta_{GLS} &= \beta_{IV} = [\underbrace{X'Z}_{(X^*)'}(Z'Z)^{-1}\underbrace{Z'X}_{X^*}]^{-1}\underbrace{X'Z}_{(X^*)'}(Z'Z)^{-1}\underbrace{Z'y}_{y^*} \qquad (5)\\ &= (X'P_zX)^{-1}X'P_zy,\end{aligned}$$

where $P_z = Z(Z'Z)^{-1}Z'$.
The variance-covariance matrix is

$$\begin{aligned}Var(\beta_{IV}) &= \sigma^2[X'Z(Z'Z)^{-1}Z'X]^{-1}\\ &= \sigma^2(X'P_zX)^{-1}\end{aligned}$$

and the distrurbance variance maybe estimated consistenly from

$$\widehat{\sigma}^2 = (y - X\beta_{IV})'(y - X\beta_{IV})/N.$$

Note that the use of N or $N-k$ or $N-l$ in the divisor does not matter asymptotically.

Special case:
When $l = k$, that is, when Z contains the same number of columns as X, we have a special case of the foregoing results.
Now $X'Z$ is $k \times k$ and nonsingular. This implies that

$$[(\underbrace{X'Z}_{(X^*)'})(Z'Z)^{-1}(\underbrace{Z'X}_{X^*})]^{-1} = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}.$$

Thus the estimator in equation (5) reduces to

$$\begin{aligned} \beta_{IV} &= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}(X'Z)(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1}Z'y \end{aligned}$$

Moreover, $Var(\beta_{IV})$ simplifies to

$$\begin{aligned} Var(\beta_{IV}) &= \sigma^2[X'Z(Z'Z)^{-1}Z'X]^{-1} \\ &= \sigma^2(Z'X)^{-1}(Z'Z)(X'Z)^{-1}. \end{aligned}$$

Two-stage least square (2SLS)

The IV estimator may also be seen as the result of a double application of least squares:

Stage (i): Regress each variable in the X matrix on Z ($X = Zd + u$) to obtain a matrix of fitted values X

$$\begin{aligned}\widehat{X} &= Z \underbrace{\delta}_{(Z'Z)^{-1}Z'X} \\ &= \underbrace{Z(Z'Z)^{-1}Z'}_{P_Z}X = P_Z X.\end{aligned}$$

Stage (ii): Regress y on $\widehat{X}$ to obtain the 2SLS estimated β vector

$$\begin{aligned}\beta_{2SLS} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \\ &= (\underbrace{X'P_Z'}_{\widehat{X}'}\underbrace{P_Z X}_{\widehat{X}})^{-1}\underbrace{X'P_Z'}_{\widehat{X}'}y \\ &= (X'P_Z X)X'P_Z y = \beta_{GLS},\end{aligned}$$

since $P_Z' P_Z = [Z(Z'Z)^{-1}Z']' = Z(Z'Z)^{-1}Z' = P_Z$ and $P_Z' P_Z = P_Z^2 = Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z' = Z(Z'Z)^{-1}Z' = P_Z$.

Thus the IV estimator can be obtained by a two-stage least-squares procedure.
The variance-covariance matrix and the estimated disturbance term are given by

$$\begin{aligned} Var(\beta_{IV}) &= \sigma^2[X'Z(Z'Z)^{-1}Z'X]^{-1} \\ &= \sigma^2(X'P_zX)^{-1} \end{aligned}$$

and

$$\widehat{\sigma}^2 = (y - X\beta_{IV})'(y - X\beta_{IV})/N.$$

Choice of instruments:

The crucial question is, where do we find the instruments?

Some of them are often variables from X matrix itself.

Any variables that are thought to be exogenous and indepedent of the disturbance are retained to serve in the Z matrix.

When some of the X variables are used as instruments, we may particion X and Z as

$$X = [X_1 \ X_2], \ Z = [X_1 \ Z_1],$$

where X_1 is of order $N \times r$ $(r < k)$, X_2 is $N \times (k - r)$, and Z_1 is $N \times (l - r)$.

It can be shown that $\widehat{X}$, the matrix of regressors in the second-stage regression, is then

$$\widehat{X} = [X_1 \ \widehat{X}_2],$$

and

$$\widehat{X}_2 = \underbrace{Z(Z'Z)^{-1}Z'}_{P_Z}X_2,$$

that is $\widehat{X}_2$ are the fitted values of X_2 obtained from the regression of X_2 on the full set of instruments: $X_2 = Z\gamma + v$
and

$$\begin{aligned} \widehat{X}_2 &= Z\widehat{\gamma}, \\ \widehat{\gamma} &= (Z'Z)^{-1}Z'X_2. \end{aligned}$$

There still remains the question of how many instruments to use.
The minimum number is k.
The asymptotic efficiency increases with the number of isntruments.
However, the small sample bias also increases with the number of instruments.
If, in fact, we select N instruments, it is simple to show that $P_Z = I$ in which case the IV estimator is simply the OLS which is biased and inconsistent.
If on the other hand, we use the minimum or close to the minimum, number of instruments, the results may also be poor.
It has been shown that the mth moment of the 2SLS estimator exists if and only if $m < l - k + 1$.
Thus, if there are just as many instruments as explanatory variables, the 2SLS estimator will not have a mean.
With one more instrument there will be a mean but not variance, and so forth.