

1 Correlated Missing Regressors

1.1 Endogeneity

Formally, the problem is that, in a model

$$Y = g(X, \beta) + \varepsilon,$$

the disturbances are *endogenous*, or equivalently, correlated with the regressors, as in

$$E[X'\varepsilon] \neq 0$$

In the Venn Diagram (Ballentine) on page 167 of Kennedy, we get a picture of this. There is a variable X which does covary with Y (red+blue+purple), but sadly some of the covariance with Y is through covariance with the error term (red). This covariance leads to bias in the OLS estimator.

1.2 Why do you get bias?

Covariation of Y with the disturbance term, which is correlated with X , is attributed to X . Consider the case where

$$Y = X\beta + \varepsilon, \varepsilon = X\Gamma + \eta \Leftrightarrow Y = X(\beta + \Gamma) + \eta.$$

The regression *loads* the response of Y to X entirely on to X . But in reality, the response of Y to X has two channels: the direct channel through β , and the indirect channel through Γ . The indirect channel is through the disturbance term: it is the derivative of ε with respect to X .

Think of the simplest possible regression model, where X is a univariate continuous variable in $[-1, 1]$ and the true coefficient is 1, but where

$$E[X'\varepsilon] = 1.$$

Here, X is positively correlated with the disturbance, which has expectation 1 when $X = 1$, and expectation -1 when $X = -1$. Draw the picture and you will see that the OLS estimator gives you the wrong slope. The reason is that it assigns variation that is due to the covariance of X and the disturbance to covariation between X and Y . Here, it will give you an estimate of 2 rather than 1.

The covariance between X and the disturbance pollutes the OLS estimator with variation that violates its identifying assumptions.

1.3 Correlated Missing Variables.

Suppose the true model is

$$\begin{aligned} Y &= X\beta + Z\Gamma + \varepsilon, \\ E \left[\begin{bmatrix} X & Z \end{bmatrix}' \varepsilon \right] &= 0 \end{aligned}$$

but we estimate a model based on

$$Y = X\beta + \varepsilon,$$

then, the expectation of the OLS estimator is

$$\begin{aligned} E[\hat{\beta}_{OLS}] &= E[(X'X)^{-1} X'(X\beta + Z\Gamma + \varepsilon)] \\ &= \beta + E[(X'X)^{-1} X'(Z\Gamma + \varepsilon)] \\ &= \beta + (X'X)^{-1} X'Z\Gamma + (X'X)^{-1} E[X'\varepsilon] \\ &= \beta + (X'X)^{-1} X'Z\Gamma \end{aligned}$$

Here, even though the ε are exogenous to the X 's, there is a bias term depending on the empirical covariance between X and Z . If $X'Z = 0$, then $\hat{\beta}_{OLS}$ is unbiased. If Z is a random variable, then if $E[X'Z] = 0$, then $\hat{\beta}_{OLS}$ is unbiased. This is why only *correlated* missing variables are a problem: uncorrelated missing variables do not induce bias.

If X and Z are correlated, then the OLS estimator is comprised of two terms added together: (1) the true coefficient on X , and (2) the marginal effect of X on $Z\Gamma$. The latter effect may be interpreted as Γ times the regression coefficients on X in a regression of Z on X .

2 The Method of Moments

A method-of-moments approach substitutes sample analogs into theoretical moments. A moment is an expectation of a power of a random variable. Sample analogs are observed variables that can be plugged in. So, every moment estimator has 3 steps:

1. Define a moment condition.
2. Plug in sample analogs for unobserved expectations.
3. Solve for estimates.

2.1 Method of Moments for OLS

For example, given the assumption that

$$E[X'\varepsilon] = 0,$$

the method-of-moments approach to estimating parameters is to substitute sample moments into the restriction:

$$X'e = 0 \Leftrightarrow X'(Y - X\beta) = 0.$$

The unique solution to this equation is the OLS estimate.

The method-of-moments approach tells you that the covariance assumption on X and the disturbances is extremely closely linked to the solution of minimising vertical distances.

2.2 Method of Moments for IV

How do you solve an endogeneity problem? Go back to the Venn Diagram. The problem is that we have red variation that OLS mistakenly attaches to X . Three general approaches are

Say you have a variable, called an instrument, usually labelled Z , that is correlated with the polluted X , but not correlated with the disturbance term. You could use that information instead of X . This is the purple area in the Venn Diagram. The purple area is not correlated with the disturbance term by construction. Thus, you can use it in an OLS regression.

The only problem is a regression of Y on the instrument gives you the marginal impact of the instrument on Y , when you really wanted the marginal impact of X on Y . You can solve this by expressing the instrument in units of X . This is the spirit of two stage least squares.

2.3 Exactly Identified Method of Moments

The moment restrictions are

$$E[Z'\varepsilon] = 0$$

If Z has K columns, then we say that the model is exactly identified. Substituting in sample moments yields

$$Z' (Y - X\hat{\beta}_{MoM}) = 0$$

Solving this for the coefficients yields:

$$\hat{\beta}_{MoM} = (Z'X)^{-1}Z'Y$$

This is sometimes referred to as *Indirect Least Squares*. Indirect Least Squares is equivalent to two-stage least squares when Z has the same rank as X :

$$\begin{aligned} \hat{X} &= P_Z X \\ \hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1} \hat{X}'Y = \\ &= (X'P_Z P_Z X)^{-1} P_Z Y = (X'P_Z X)^{-1} X'P_Z Y \\ &= (X'Z (Z'Z)^{-1} Z'X)^{-1} X'Z (Z'Z)^{-1} Z'Y \\ &= (Z'X)^{-1} (Z'Z) (X'Z)^{-1} X'Z (Z'Z)^{-1} Z'Y \\ &= (Z'X)^{-1} Z'Y, \end{aligned}$$

because with square matrices, you can rearrange within the inverse.

Consider a one-dimensional X and Z , and consider regressing Y on Z and

X on Z :

$$\begin{aligned}\widehat{\beta}_{OLS}^{YonZ} &= (Z'Z)^{-1}Z'Y \\ \widehat{\beta}_{OLS}^{XonZ} &= (Z'Z)^{-1}Z'X \\ \widehat{\beta}_{2SLS}^{YonX} &= \widehat{\beta}_{MM}^{YonX} = \frac{\widehat{\beta}_{OLS}^{YonZ}}{\widehat{\beta}_{OLS}^{XonZ}}\end{aligned}$$

Here, we see that the IV estimator (2SLS, MM) is given by the ratio of two OLS estimators. It is the ratio of the derivatives of Y and X with respect to Z , and for that reason, we think of it as indirectly illuminating the derivative of Y with respect to X .

Thus, regressing Y on Z looks the same as 2SLS of Y on X in the Ballentines, but the Ballentine misses out the rescaling shown above.

2.4 Generalised Method of Moments

If Z contains more columns than X , we say that the model is *overidentified*. In this case, you have too much information, and you can't get the moment restriction all the way to zero. Instead, you solve for the coefficients by minimising a quadratic form of the moment restriction. This is called *Generalised Method-of-Moments* (GMM). The GMM estimator is exactly equal to the two stage least squares estimator in linear homoskedastic models.

Assume Z has $J > K$ columns. GMM proposes a criterion "get $Z'e$ as close to zero as you can by choice of the parameter vector β ". Since β only has K elements and $Z'e$ has more than K elements, you can't generally get $Z'e$ all the way to zero. So, you minimize a quadratic form in $Z'e$ with some kind of weighting matrix in the middle. Consider

$$\min_{\beta} e'Z\Omega^{-1}Z'e$$

where Ω is a matrix that puts weight on the bits of $Z'e$ that you think are most important, or most informative about $Z'e$.

Since $E[Z'\varepsilon] = 0$, it doesn't matter asymptotically how you weight the various bits of it. You could use $\Omega = I_J$ if you wanted, and the estimator would be consistent. Minimisation with $\Omega = I_J$ is often referred to as *Minimum Distance Estimation*. The point here is that **any choice of Ω** yields a consistent estimator. However, different choices of Ω yield estimators of different efficiency.

A natural choice of Ω is the covariance matrix of $Z'\varepsilon$. By assumption, $E[Z'\varepsilon] = 0$, so its covariance $E[Z'\varepsilon\varepsilon'Z]$ is equal to its mean squared error. If an element of $Z'\varepsilon$ is always really close to zero, then you'd want to pay a lot of attention to keeping this element close to zero. In contrast, if an element of $Z'\varepsilon$ varied wildly, you wouldn't want to pay too much attention to it in choosing your β . Since the weighting matrix is Ω^{-1} , elements of $Z'\varepsilon$ that are wildly varying, and have big variance, get small weight in the quadratic form; elements which are always close to zero, and have small variance, get big weight in the quadratic form.

Hansen (1982) shows that if Ω is the covariance matrix of $Z'\varepsilon$, then the GMM estimator is asymptotically efficient.

Given homoskedasticity,

$$V[Z'\varepsilon] = E[Z'\varepsilon\varepsilon'Z] = E[Z'\sigma^2 I_N'Z] = \sigma^2 Z'Z$$

So, we minimize

$$\begin{aligned} \min_{\beta} (Y - X\beta)' Z \frac{1}{\sigma^2} (Z'Z)^{-1} Z' (Y - X\beta) &= \\ \min_{\beta} (Y - X\beta)' Z (Z'Z)^{-1} Z' (Y - X\beta) &= \\ \min_{\beta} (Y - X\beta)' P_Z (Y - X\beta) &= \\ \min_{\beta} (Y - X\beta)' P_Z' P_Z (Y - X\beta) &= \end{aligned}$$

yielding a first-order condition

$$\begin{aligned} 2X'P_Z' (Y - X\beta) &= 0 \\ \Leftrightarrow \\ X^{1rststage'} (Y - X\beta) &= 0. \end{aligned}$$

That is, choose β to make e orthogonal to \hat{X} .

That is, given homoskedasticity, the GMM estimator is the 2SLS estimator:

$$\begin{aligned} X'P_Z' (Y - X\beta) &= 0 \\ X'P_Z' Y &= X'P_Z' X\beta \\ \hat{\beta}_{GMM} = \hat{\beta}_{2SLS} &= (X'P_Z' X)^{-1} X'P_Z' Y \end{aligned}$$

Since all exogenous variables correspond to columns of X that are also in Z , and defining as an instrument as an exogenous variable that is in Z but not in X , this implies that you need at least one instrument in Z for every endogenous variable in X . Consider

$$X = [X_1 \ X_2], \quad Z = [X_1 \ Z_2],$$

a case where a subvector of X is exogenous (because it shows up in Z), and the rest is endogenous. Here, X_2 is endogenous.

$$Y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

The two stage approach to IV is to estimate X on Z and use its predicted values on the RHS instead of X .

What are its predicted values? Since Z contains the exogenous pieces of X , this is equivalent to regressing

$$Y = X_1\beta_1 + \hat{X}_2\beta_2 + \varepsilon$$

where

$$\hat{X}_2 = Z (Z'Z)^{-1} Z' X_2$$

the predicted values from a regression of the endogenous X 's on the entire set of Z .