q5

rafehq

November 2018

1 Q6

a

Suppose we have that $\mathbb{E}(U) = \alpha \neq 0$. Then, we have

$$E(U) = E(log(Y) - \beta_0 - \beta_1 X_1 - \beta_2 X_2) = \alpha$$

But, then we can write

$$log(Y) = \tilde{\beta_0} + \beta_1 X_1 + \beta_2 X_2 + \tilde{U}$$

, wherein $\tilde{\beta}_0 = \beta_0 + \alpha$ and $\tilde{U} = U - \alpha$. Then, we see that

$$\begin{aligned} E(\tilde{U}) &= E(log(Y) - \tilde{\beta}_0 - \beta_1 X_1 - \beta_2 X_2) \\ &= E(log(Y) - \beta_0 - \alpha - \beta_1 X_1 - \beta_2 X_2) \\ &= E(log(Y) - \beta_0 - \beta_1 X_1 - \beta_2 X_2) - E(\alpha) \\ &= \alpha - \alpha = 0 \end{aligned}$$

Note, as this states, we cannot separate $E(U)$ from $E(\beta_0 + U)$, because U is not a variable included in the regression. Namely, the mean of U will be included in the expectation of the calculated constant in the regression.

b

We can say X_k is exogenous if $E(X_k U) = 0$. This means X_k is orthogonal to the error term. With our examples from class, we can say that there is no measurement error in X_k, there are no variables omitted in the regression that are correlated with X_k and X_k is not determined simultaneously with Y.
Similarly, we can say X_k is endogenous if $E(X_k U) \neq 0$.
In our particular example, we would guess that $E(UX_1) \neq 0$ as there are variables not included in the regression that are correlated with years of education and are also correlated with Y, but are not included in our regression. The

classic example of this is *ability.* An individual's ability, such as her intelligence and sedulousness is likely to influence her decision to enroll in more schooling and are also likely to cause her to have a higher hourly wage. Without controlling for ability however, we are likely to attribute to years of education what is partially caused by ability.

c

We say the instrument is exogenous if it is uncorrelated with the error term; $E(Zu) = 0$.
We say the instrument is relevant if the rank$(\mathbb{E}(ZX') = k + 1$.

d

We can use the IV estimator; namely

$$\hat{\beta}_{IV} = (\frac{1}{N}\sum_{i=1}^{N} Z_i X_i')^{-1}(\frac{1}{N}\sum_{i=1}^{N} Z_i Y_i')$$

In the exactly, identified case, this would be equivalent to the 2SLS estimator. However, for the overidentified case (where we have more than one instrumental variable for X_1), we go over the 2SLS procedure. We would go about

1. Regressing X_1 on the entire matrix Z, and we can call the estimated coefficients π.
2. Getting the estimated X_1 from the regression in (1). Namely, $\hat{X}_1 = \hat{\pi}'Z$.
3. Regressing Y on $\hat{X}_1$ and X_2. This will give us

$$\hat{\beta}_{TSLS} = (\frac{1}{N}\sum_{i=1}^{N} \hat{\pi}' Z_i Z_i' \hat{\pi})^{-1}(\frac{1}{N}\sum_{i=1}^{N} \hat{\pi}' Z_i Y_i)$$

e

We can look at the regression of

$$X_1 = \gamma_0 + \gamma_1 Z_1 + \gamma_3 X_2 + \epsilon$$

Here, clearly, the best linear predictor of X_1 given Z is $\gamma_0 + \gamma_1 Z_1 + \gamma_3 X_2$. We would like to test if Z_1 and X_1 are uncorrelated; namely we would like to test the null hypothesis $\gamma_1 = 0$ against the alternative hypothesis $\gamma_1 \neq 0$.
We recall that for the vector of coefficients γ,

$$\sqrt{N}(\hat{\gamma} - \gamma) \to N(0, \Omega)$$

We can now build the estimator $\hat{\Omega}$ through the residuals calculate the γ_1 through OLS to reject the null at α if:

$$|\frac{\sqrt{N}(\hat{\gamma}_1)}{\sqrt{\Omega_{2,2}}}| > z_{1-\frac{\alpha}{2}}$$

f

Note, as $Z = (1, Z_1, X_2)'$, we see that the rank condition would be violated if we tried using $Z_1 = X_2$ as the instrument (we are adding fewer instrumental variables to the model here than the number of endogenous variables we have).

Now, suppose, alternatively, that we remove X_2 from the regression and instead try to estimate

$$log(Y) = \beta_0 + \beta_1 X_1 + \tilde{U}$$

Then, we may no longer have that $Cov(Z, \tilde{U}) = 0$. Indeed, if we are right about the model and $\beta_2 \neq 0$, this would have to be the case.