Homework 3

Tate Mason

An ECON - 8070 Homework Assignment

November 11, 2024

Question 4.2

Problem

(a)

Show that, under random sampling and the zero conditional mean assumption $\mathbb{E}(u|\mathbf{x}) = 0$, $\mathbb{E}(\hat{\beta}|\mathbf{X}) = \beta$ if $\mathbf{XX'}$ is nonsingular. (Hint: use property CE.5 in the appendix of chapter 2)

(b)

In addition to the assumption from part a, assume that $\text{var}(u|\mathbf{x}) = \sigma^2$. Show that $\text{var}(\hat{\beta}|\mathbf{X}) = \sigma^2((\mathbf{XX'})^{-1})$

Solutions

(a)

Proof. The OLS estimator is given by $\hat{\beta} = (X'X)^{-1}X'y$ and the linear model gives us that $y = X\beta + u$. If we substitute that into the formula for $\hat{\beta}$, we get $\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$. Next, taking the conditional expectation of $\hat{\beta}$ given X and using the zero mean assumption, we can show that $\mathbb{E}(\hat{\beta}|X) = (X'X)^{-1}X\beta + (X'X)^{-1}X'0 \rightarrow \mathbb{E}(\hat{\beta}|X) = (X'X)^{-1}X'X\beta \rightarrow \mathbb{E}(\hat{\beta}|X) = \beta$ □

(b)

Proof. The formula for the variance of the OLS estimator $\hat{\beta} = \mathbb{E}[(\hat{\beta} - \mathbb{E}(\beta\hat{|}X))(\hat{\beta} - \mathbb{E}(\hat{\beta}|X))'|X']$. Using the result from (a), we can say that $var(\hat{\beta}|X) = \mathbb{E}[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'|X] = (X'X)^{-1}X'\mathbb{E}(uu'|X)X(X'X)^{-1}$. As the problem tells us that $\text{var}(u|X) = \sigma^2$, we can state $\mathbb{E}(uu'|X) = \sigma^2 I$. Thus, $\text{var}(\hat{\beta}|X) = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$ □

Question 4.3

Problem

Suppose that in the linear model 4.5, $\mathbb{E}(\mathbf{x'}u) = \mathbf{0}$ (where $\mathbf{x}$ contains unity), $\text{var}(u|\mathbf{x}) = \sigma^2$, but $\mathbb{E}(u|\mathbf{x}) \neq \mathbb{E}(u)$.

(a)

Is it true that $\mathbb{E}(u^2|\mathbf{x}) = \sigma^2$?

(b)

What relevance does part a have for OLS estimation?

Solutions

(a)

No, this is not true.

Proof. $(\text{var}u|x) = \mathbb{E}(u^2|x) - \mathbb{E}(u|x)^2$. From the question, we can say that $\sigma^2 = \mathbb{E}(u^2|x) - \mathbb{E}(u|x)^2$. Due to the fact that $\mathbb{E}(u|x) \neq \mathbb{E}(u)$, we know that u varies with x, and the same with $\mathbb{E}(u|x)^2$. So, rearranging the equation, $\mathbb{E}(u^2|x) = \sigma^2 + \mathbb{E}(u|x)^2$. Since, as stated, $\mathbb{E}(u|x)^2$ varies with x, so too must $\mathbb{E}(u^2|x)$. With that conclusion we can say that $\mathbb{E}(u^2|x) \neq \sigma^2$ □

(b)

When heteroskedasticity is present, the usual OLS S.E. will be incorrect due to the assumption of homoskedasticity. Thus, robust S.E. will be needed to produce meaningful estimation with OLS.

Question 4.17

Problem

Consider the standard linear model $y = \boldsymbol{x\beta} + \boldsymbol{u}$ under Assumptions OLS.1 and OLS.2. Define $h(\boldsymbol{x}) \equiv \mathbb{E}[u^2|\boldsymbol{x}]$. Let $\hat{\boldsymbol{\beta}}$ be the OLS estimator, and show that we can always write:

$$\text{Avar}\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left[\mathbb{E}[\boldsymbol{x}'\boldsymbol{x}]\right]^{-1} \mathbb{E}[h(\boldsymbol{x})\boldsymbol{x}'\boldsymbol{x}] \left[\mathbb{E}[\boldsymbol{x}'\boldsymbol{x}]\right]^{-1} \tag{1}$$

This expression is useful when $\mathbb{E}[\boldsymbol{u}|\boldsymbol{x}] = \boldsymbol{0}$ for comparing the asymptotic variances of OLS and weighted least squares estimators; see, for example, Wooldridge (1994b).

Solution

Proof. □

Question 5.1

Problem

In this problem you are to establish the algebraic equivalence between 2SLS and OLS estimation of an equation containing an additional regressor. Although the result is completely general, for simplicity consider a model with a single (suspected) endogenous variable:

$$\begin{aligned} y_1 &= z_1\delta_1 + \alpha_1 y_2 + u_1, \\ y_2 &= z\pi_2 + v_2. \end{aligned}$$

For notational clarity, we use y_2 as the suspected endogenous variable and z as the vector of all exogenous variables. The second equation is the reduced form for y_2. Assume that z has at least one more element than z_1. We know that one estimator of (δ_1, α_1) is the 2SLS estimator using instruments x. Consider an alternative estimator of (δ_1, α_1): (a) estimate the reduced form by OLS, and save the residuals $\hat{v}_2$; (b) estimate the following equation by OLS:

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + \rho\hat{v}_2 + \text{error}. \tag{5.52}$$

Show that the OLS estimates of δ_1 and α_1 from this regression are identical to the 2SLS estimators. (Hint: Use the partitioned regression algebra of OLS. In particular, if $\hat{y} = x_1\beta_1 + x_2\beta_2$ is an OLS regression, β_1 can be obtained by first regressing x_1 on x_2, getting the residuals, say $\ddot{x}_1$, and then regressing y on $\ddot{x}_1$; see, for example, Davidson and MacKinnon (1993, Section 1.4). You must also use the fact that z_1 and $\hat{v}_2$ are orthogonal in the sample.)

Solution

Proof. 2SLS in this case would work by first regressing y_2 on z to get $\hat{y}_2$. Then, we would regress y_1 on z_1 and $\hat{y}_2$. For the alternative method, $\hat{v}_2 = y_2 - z\hat{\pi}_2$ which gives us the residuals from the reduced form problem. Then, we would run OLS on equation 5.52. To derive δ_1 and α_1, we regress y_2 and z_1 on $\hat{v}_2$ and get the residuals. Then, we regress y_1 on the residuals from the last step. Since, as given, $\hat{v}_2$ is orthogonal to z_1, the residual from the regression of z_1 on $\hat{v}_2$ is z_1. The residual from the regression of y_2 on $\hat{v}_2$ is $\hat{y}_2$. This is the same result as when regressing y_1 on z_1 and $\hat{y}_2$. The two estimates, therefore, show identical estimates for δ_1 and α_1. □

Question 5.2

Problem

Consider a model for the health of an individual:

$$\begin{aligned} \text{health} = \beta_0 + \beta_1\text{age} + \beta_2\text{weight} + \beta_3\text{height} \\ +\beta_4\text{male} + \beta_5\text{work} + \beta_6\text{exercise} + u_1, \end{aligned} \tag{2}$$

where *health* is some quantitative measure of the person's health; *age*, *weight*, *height*, and *male* are self-explanatory; *work* is weekly hours worked; and *exercise* is the hours of exercise per week.

Parts

(a)

Why might you be concerned about *exercise* being correlated with the error term u_1?

(b)

Suppose you can collect data on two additional variables, *disthome* and *distwork*, the distances from home and from work to the nearest health club or gym. Discuss whether these are likely to be uncorrelated with u_1.

(c)

Now assume that *disthome* and *distwork* are in fact uncorrelated with u_1, as are all variables in equation (2) with the exception of *exercise*. Write down the reduced form for *exercise*, and state the conditions under which the parameters of equation (2) are identified.

(d)

How can the identification assumption in part c be tested?

Solutions

(a)

Things like health conditions can affect ability to exercise as well as those who care more about health will be more likely to exercise and be healthier. These unobserved factors, as well as the potential for bad health to effect ability exercise, are reasons why exercise may be correlated with the error term.

(b)

These are likely uncorrelated factors as the choice of a dwelling or a workplace is typically decided by more factors than just proximity to a gym. Thus, it would be a valid addition to the model.

(c)

The reduced form equation is as follows:

$$exercise = \pi_0 + \pi_1 age + \pi_2 weight + \pi_3 height + \pi_4 male + \pi_5 work + \pi_6 disthome + \pi_7 distwork + v_2$$

Identification has three primary conditions:

a) Rank condition states that one of π_6, π_7 must be non-zero

b) The exclusion restriction is assumed in this question

c) Overidentified equation as the instruments outnumber endogenous variables

(d)

One example of a test for the identification assumption in (c) is the first stage F-test. Tests joint significance of instruments in reduced form.

Question 5.11

Problem

A model with a single endogenous explanatory variable can be written as

$$y_1 = z_1\delta_1 + \alpha_1 y_2 + u_1, \tag{3}$$
$$E(z'u_1) = 0, \tag{4}$$

where $z = (z_1, z_2)$. Consider the following two-step method, intended to mimic 2SLS:

(a)

Regress y_2 on z_2, and obtain fitted values, $\hat{y}_2$. (That is, z_1 is omitted from the first-stage regression.)

(b)

Regress y_1 on z_1, $\hat{y}_2$ to obtain $\hat{\delta}_1$ and $\hat{\alpha}_1$. Show that $\hat{\delta}_1$ and $\hat{\alpha}_1$ are generally inconsistent. When would $\hat{\delta}_1$ and $\hat{\alpha}_1$ be consistent? (Hint: Let y_2^0 be the population linear projection of y_2 on z_2, and let a_2 be the projection error: $y_2^0 = z_2\lambda_2 + a_2$, $E(z'a_2) = 0$. For simplicity, pretend that λ_2 is known rather than estimated; that is, assume that $\hat{y}_2$ is actually y_2^0. Then, write)

$$y_1 = z_1\delta_1 + \alpha_1 y_2^0 + \alpha_1 a_2 + u_1 \tag{5}$$

and check whether the composite error $\alpha_1 a_2 + u_1$ is uncorrelated with the explanatory variables.

Solution

Proof. As given in the hint, $y_2^0 = z_2\lambda_2 + a_2$ such that $\mathbb{E}(z_2'a_2) = 0$. Substituting this into the main equation, $y_1 = z_1\delta_1 + \alpha_1(z_2\lambda_2 + a_2) + u_1 y_1 = z_1\delta_1 + \alpha_1 z_2\lambda_2 + (\alpha_1 a_2 + u_1)$. For consistency, $\alpha_1 a_2 + u_1$ needs to be uncorrelated with both z_1 and y_2^0. From assumptions given in the question, we know that $\mathbb{E}(z_1'u_1) = 0$. However, $\mathbb{E}(z_1'a_2) \neq 0$ due to the exlusion of z_1 from the linear projection. With this in my mind, we can say that $\hat{\delta}_1, \hat{\alpha}_1$ are generally inconsistent. They would be consistent if z_1, z_2 are uncorrelated, y_2 is unrelated to z_1, or if $\alpha_1 = 0$. Important takeaways are that due to the omission of z_1 from the first stage is akin to incorrectly excluding relevant instruments, creating correlation between a_2 and the exogenous variables. Thus, it is necessary to use valid instruments in the first stage. □