Assignment 2

630516am and 590049bm

23 Nov 2021

Question 1

```
set.seed(2021)
```

```
b0 <- 3
b1 <- 5
b2 <- 8

gamma <- 1

x1 <- rnorm(5000, mean = 1, sd = 1)
x2 <- rnorm(5000, mean = 2, sd = 1)

z <- rgamma(5000, shape = 1.2, scale = 1.1)

sigma_sq <- 1*exp(gamma*z)
epsilon <- rnorm(5000, mean = 0, sd = sqrt(sigma_sq))

y <- b0 + b1*x1 + b2*x2 + epsilon
```

```
model1 <- lm(y ~ x1 + x2)

modelsummary(model1,
             vcov = c("iid", "HC0"),
             gof_map = gm,
             stars = T)
```

```
lmtest::bptest(formula = y ~ x1 + x2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  y ~ x1 + x2
```

	Model 1	Model 2
(Intercept)	2.609***	2.609***
	(0.375)	(0.315)
x1	5.009***	5.009***
	(0.149)	(0.078)
x2	8.080***	8.080***
	(0.152)	(0.097)
N	5000	5000
Adj. R2	0.44	0.44

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
## BP = 0.50429, df = 2, p-value = 0.7771
```

```r
gamma <- 0

sigma_sq <- 1*exp(gamma*z)
epsilon <- rnorm(5000, mean = 0, sd = sqrt(sigma_sq))

y <- b0 + b1*x1 + b2*x2 + epsilon
```

```r
model1 <- lm(y ~ x1 + x2)
model2 <- lm(y ~ x1 + x2, weights = sigma_sq)
```

Question 2

1. Show that the OLS estimator of the parameter β is not consistent.
2. Derive plim (b) where b is the OLS estimator of β. Determine the sign of the magnitude of the inconsistency when $0 < \beta < 1$, that is, the sign of plim(b) - β when $0 < \beta < 1$.

First, we demean the two variables so that the constant-term α equals zero. Then we regress $\tilde{C} = \beta\tilde{D} + \epsilon$. We can do this because of Frisch-Waugh-Lovell. The estimate that we get is:

$$\hat{\beta} = (\tilde{D}^T\tilde{D})^{-1}\tilde{D}^T C_{=}(\tilde{D}^T\tilde{D})^{-1}(\beta\tilde{D} + \epsilon)$$

and

$$\mathbb{E}[\hat{\beta}] = \beta + (\tilde{D}^T\tilde{D})^{-1}\tilde{D}^T\epsilon$$

Evaluating the probability limit gives:

$$\text{plim}_{n\to\infty}(\hat{\beta}) = \beta + \text{plim}(\frac{1}{n}\tilde{D}^T\tilde{D})^{-1} \cdot \text{plim}(\frac{1}{n}\tilde{D}^T\epsilon)$$

which simplifies to:

$$\beta + \frac{1}{\text{Var}(D)} \cdot \frac{1}{1-\beta}\sigma^2$$

by the fact that variances and covariances are the same after demeaning, and by the reduced form equation for D made explicit below. Under $0 < \beta < 1$, since variances are positives, the right term can only be positive and thus the bias is always positive.

Substituting equation (2) into equation (1) and solving for C gives:

$$C = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta}Z_i + \frac{1}{1-\beta}\epsilon_i$$

substituting this back in the definition for D gives:

$$D = \frac{\alpha}{1-\beta} + \left(\frac{\beta}{1-\beta} + 1\right)Z_i + \frac{1}{1-\beta}\epsilon_i$$

From this, we can calculate Cov(D, ϵ_i), which is $\frac{1}{1-\beta}\text{Var}(\epsilon) = \frac{1}{1-\beta}\sigma^2$.

3. Find an instrumental variable (IV) for the endogenous variable D i and argue why it could be an IV.

The instrumental variable could be Z, because it is relevant, i.e. Cov(D, Z) $\neq 0$. Also, it is exogenous (valid), as it is exogenously generated and has no correlation with the error term ϵ according to the DGP sketched out here.

4. Derive b_{IV}, the IV estimator of β in terms of the variables C, D, and Z step by step.

First, suppose X is a matrix consisting of a column of 1's and D, so we can write:

$$\mathbb{E}[z_i \epsilon_i] = 0 = \frac{1}{n} \sum z_i (c_i - \alpha - \beta D_i) = \frac{1}{n} \sum z_i (c_i - x_i \beta)$$

Using this moment condition to solve for β, we retrieve the b_{IV} estimator:

$$\hat{b}_{IV} = \left(\sum z_i^T x_i \right)^{-1} \left(z_i^T c_i \right) = (Z^T X)^{-1} Z^T C$$

5. Use the expression of b_{IV} to show that it is consistent.

$$b_{IV} = (Z^T X)^{-1} Z^T C = (Z^T X)^{-1} Z^T (X\beta + \epsilon) = (Z^T X)^{-1} Z^T X \beta + (Z^T X)^{-1} Z^T \epsilon = \beta + (Z^T X)^{-1} Z^T \epsilon$$

Evaluating the plim of this estimator then gives:

$$\text{plim}(b_{IV}) = \beta + \text{plim}(Z^T X)^{-1} \cdot \text{plim}(Z^T \epsilon)$$

where the last factor goes to zero as $n \to \infty$.

Question 3

1. Plot the distribution of the growth rate of employment and of import exposure 1990-2007 across US commuting zones