

Problem Set #2

Danny Edgel
Econ 709: Economic Statistics and Econometrics I
Fall 2020

November 21, 2020

Collaborated with Sarah Bass, Emily Case, Michael Nattinger, and Alex Von Hafften

Question 1

Recall that $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$. Then, if $Z = XC$,

$$\begin{aligned}\hat{\beta}_Z &= (Z'Z)^{-1}Zy = [(XC)'XC]^{-1}(XC)'y \\ &= (C'X'XC)^{-1}C'X'y = C^{-1}(X'X)^{-1}C'^{-1}C'X'y \\ &= C^{-1}(X'X)^{-1}X'y = C^{-1}\hat{\beta}_{OLS}\end{aligned}$$

Also recall that $\hat{\varepsilon}_{OLS} = Y - X\hat{\beta}_{OLS}$. Then,

$$\begin{aligned}\hat{\varepsilon}_Z &= Y - Z\hat{\beta}_Z = y - Z(Z'Z)^{-1}Zy \\ &= (I - XC((XC)'XC)^{-1}XC)y = (I - XCC^{-1}(X'X)^{-1}C'^{-1}C'X)y \\ &= (I - X(X'X)^{-1}X)y = y - X(X'X)^{-1}Xy \\ &= y - X\hat{\beta}_{OLS} = \hat{\varepsilon}_{OLS}\end{aligned}$$

Question 2

3.5) Recall from question 1 that $\hat{\varepsilon}_{OLS} = (I - X(X'X)^{-1}X')Y$. Then,

$$\begin{aligned}\hat{\beta}_e &= (X'X)^{-1}X'\hat{\varepsilon}_{OLS} = (X'X)^{-1}X'(I - X(X'X)^{-1}X')Y \\ &= ((X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X')Y = ((X'X)^{-1}X' - (X'X)^{-1}X')Y \\ &= 0\end{aligned}$$

3.6) Let $\hat{Y} = X(X'X)^{-1}X'Y$ and $\hat{\beta}_Y$ represent the OLS coefficient from a regression of \hat{Y} on X . Then,

$$\begin{aligned}\hat{\beta}_Y &= (X'X)^{-1}X'\hat{Y} = (X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'Y = \hat{\beta}_{OLS}\end{aligned}$$

- 3.7) Let $X = [X_1 \ X_2]$ be an m by n matrix and recall that $P = X(X'X)^{-1}X'$ and $M = I - P$. Let $n = n_1 + n_2$, where X_1 is an m by n_1 matrix and X_2 is m by n_2 . Then, we can define $\Gamma = \begin{pmatrix} I_{n_1} \\ 0 \end{pmatrix}$ such that $X_1 = X\Gamma$. Thus,

$$\begin{aligned} PX_1 &= PX\Gamma = X(X'X)^{-1}X'X\Gamma = X\Gamma = X_1 \\ MX_1 &= (I - P)X_1 = X_1 - PX_1 = X_1 - X_1 = 0 \end{aligned}$$

Question 3

- 3.11) Let X contain only a non-zero constant, $c \in \mathbb{R}$, such that $X = c\mathbb{1}_n$, where n is the number of elements in Y and $\mathbb{1}_n$ is an $n \times 1$ vector of ones. Then,

$$\begin{aligned} \hat{Y} &= X(X'X)^{-1}X'Y = (c\mathbb{1}_n)[(c\mathbb{1}_n)'(c\mathbb{1}_n)]^{-1}(c\mathbb{1}_n)'Y \\ &= c\mathbb{1}_n(c^2(\mathbb{1}_n'\mathbb{1}_n))^{-1}c(\mathbb{1}_n'Y) = c^2\mathbb{1}_n(c^2n)^{-1}n\bar{Y} \\ &= \frac{c^2n}{c^2n}\bar{Y}\mathbb{1}_n = \bar{Y}\mathbb{1}_n \end{aligned}$$

Thus, \hat{Y} is a column vector where every entry is \bar{Y}

- 3.12) Equation (3.53) cannot be estimated by OLS. Equation (3.53) can be rewritten as $Y = X\beta + \varepsilon$, where $X = [\mathbb{1}_n \ D_1 \ D_2]$ and $\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$. $D_1 + D_2 = \mathbb{1}_n$, so $\text{rank}(X) \neq k$, violating the first Gauss-Markov assumption.

- (a) Neither (3.54) nor (3.55) is more general. The two specifications have the same explanatory power. In (3.54), the average of Y for men is given by α_1 , and the average for women is given by α_2 . In (3.55), the averages are $\mu + \phi$ and μ , respectively. Thus, given the parameters for one specification, you could calculate the parameters of the other with:

$$\begin{aligned} \mu + \phi &= \alpha_1 & \phi &= \alpha_2 - \alpha_1 \\ \mu &= \alpha_2 & \alpha_2 &= \mu \end{aligned}$$

- (b) $\mathbb{1}_n' D_1 = n_1$, $\mathbb{1}_n' D_2 = n_2$

- 3.13) (a) Letting $X = [D_1 \ D_2]$ and $\hat{\beta} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix}$, we can solve:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y \\ &= \begin{pmatrix} \mathbb{1}_n' D_1 & 0 \\ 0 & \mathbb{1}_n' D_2 \end{pmatrix}^{-1} \begin{pmatrix} D_1' Y \\ D_2' Y \end{pmatrix} = \frac{1}{n_1 n_2} \begin{pmatrix} n_2 & 0 \\ 0 & n_1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n D_{1i} Y_i \\ \sum_{i=1}^n D_{2i} Y_i \end{pmatrix} \\ &= \frac{1}{n_1 n_2} \begin{pmatrix} n_2 \sum_{i=1}^n D_{1i} Y_i \\ n_1 \sum_{i=1}^n D_{2i} Y_i \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} \sum_{i=1}^n D_{1i} Y_i \\ \frac{1}{n_2} \sum_{i=1}^n D_{2i} Y_i \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix} \end{aligned}$$

- (b) In plain English, Y^* is the demeaned value of Y , using the means for men and women separately. Econometrically, as shown below, Y^* is the residualized value of Y , or, rather, the value of Y that cannot be explained by gender alone:

$$\begin{aligned} Y^* &= Y - D_1 \bar{Y}_1 - D_2 \bar{Y}_2 = Y - (D_1 \bar{Y}_1 + D_2 \bar{Y}_2) \\ &= Y - [D_1 \ D_2] \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \hat{\mu} \end{aligned}$$

It logically follows, then, that X^* is the residualized value of X , from a regression of X on D_1 and D_2 .

Question 4

3.16

Question 5

3.21)

3.22)

3.23)

Question 6

(due w/ PS3)

3.24)

3.25)

Question 7

Given the $n \times 1$ vector y and the $n \times k$ matrix X , assume $\text{Prank}(X) = k$; $\mathbb{E}(y|X) = X\beta$; and $\text{Var}(y|X) = \sigma^2 I$.

Partiion X : $X = [X_1 \ X_2]$ where X_1 is $n \times k_1$, X_2 is $n \times k_2$, and $k_1 + k_2 = k$.

Similarly partition β : $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, where β_1 is $k_1 \times 1$ and β_2 is $k_2 \times 1$.

- Consider the OLS regression of y on X that yields the OLS estimator $\hat{\beta}$. What is $\mathbb{E}(\hat{\beta}_1|X)$? Simplify your answer.**
- Let $\hat{y} = X\hat{\beta}$. Now, consider the OLS regression of \hat{y} on X_1 that yields the OLS estimator $\hat{\beta}_1$. What is $\mathbb{E}(\hat{\beta}_1|X)$? (Simplify your answer.) Is $\hat{\beta}_1$ an unbiased estimator of β_1 ?**

- (c) Consider the OLS regression of y on X_1 that yields the OLS estimator β_1 . Let $\tilde{y} = X_1\tilde{\beta}_1$. Now consider the OLS regression of \tilde{y} on X that yields the OLS estimator $\tilde{\beta}$. How is $\tilde{\beta}$ related to $\tilde{\beta}_1$? (Provide a mapping between $\tilde{\beta}$ and $\tilde{\beta}_1$ that does not involve X .)
- (d) What is the R^2 for the OLS regression of \tilde{y} on X (from part (c))? Simplify your answer.
- (e) What is $Var(\tilde{\beta}_1|X)$? Simplify your answer.