Problem Set #2 (2^{nd} Half) Solutions

Economics 709 **Fall 2020**

3.2

$$\begin{aligned}\tilde{\beta} &= \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{y} \\ &= \left(\boldsymbol{C}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{C}\right)^{-1}\boldsymbol{C}'\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{C}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{C}'^{-1}\boldsymbol{C}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{C}^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{C}^{-1}\hat{\beta}.\end{aligned}$$

$$\tilde{e} = \mathbf{y} - \mathbf{Z}\,\tilde{\beta} = \mathbf{y} - (\mathbf{XC})\boldsymbol{C}^{-1}\,\hat{\beta} = \mathbf{y} - \mathbf{X}\,\hat{\beta} = \hat{e}$$

All column vectors of $\boldsymbol{Z}$ are a linear combination of the column vectors of $\boldsymbol{X}$. Thus the column space of $\boldsymbol{Z}$ is the same as the column space of $\boldsymbol{X}$ with the same dimension k. Thus fitted values and residuals from regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ are the same as those from the regression of $\boldsymbol{y}$ on $\boldsymbol{Z}$. The OLS estimates based on $\mathbf{Z}$ are a linear combination of the OLS estimates based on $\mathbf{X}$ that inverts the transformation from $\mathbf{X}$ to $\mathbf{Z}$.

3.5 The OLS coefficient from regression of $\hat{\boldsymbol{e}}$ on $\boldsymbol{X}$ is,

$$\tilde{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{e}} = 0,$$

since $\boldsymbol{X}'\hat{\boldsymbol{e}} = \boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{y} = 0$.

3.6 Let $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Note that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$, so $\boldsymbol{X}'\boldsymbol{P} = \boldsymbol{X}'\boldsymbol{P}' = (\boldsymbol{P}\boldsymbol{X})' = \boldsymbol{X}'$, by symmetry of $\boldsymbol{P}$. The OLS coefficient from a regression of $\hat{y}$ on $\boldsymbol{X}$ is,

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{y} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{P}y = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'y = \hat{\beta}.$$

3.7 Note that $\boldsymbol{P}\boldsymbol{X} = \boldsymbol{X}$, $\boldsymbol{M}\boldsymbol{X} = 0$. By writing out the partitioned matrix, $[\boldsymbol{X_1}\ \boldsymbol{X_2}] = \boldsymbol{X} = \boldsymbol{P}\boldsymbol{X} = [\boldsymbol{P}\boldsymbol{X_1}\ \boldsymbol{P}\boldsymbol{X_2}]$, and $\boldsymbol{M}\boldsymbol{X} = [\boldsymbol{M}\boldsymbol{X_1}\ \boldsymbol{M}\boldsymbol{X_2}]$. Therefore, $\boldsymbol{P}\boldsymbol{X_1} = \boldsymbol{X_1}$, and $\boldsymbol{M}\boldsymbol{X_1} = 0$.

3.11 Note that $\hat{\boldsymbol{Y}}_i = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}$, so $\hat{\boldsymbol{e}}_i = \boldsymbol{Y}_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} = \boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_i$ and $\hat{\boldsymbol{Y}}_i = \boldsymbol{Y}_i - \hat{\boldsymbol{e}}_i$. Since $\boldsymbol{X}'\hat{\boldsymbol{e}} = \boldsymbol{0}$, when $\boldsymbol{X}$ contains a constant $\boldsymbol{0} = \boldsymbol{1}'\hat{\boldsymbol{e}} = \sum_i \hat{\boldsymbol{e}}_i$. Then,

$$\frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_i \hat{\boldsymbol{Y}}_i = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_i \boldsymbol{Y}_i - \frac{\boldsymbol{1}}{\boldsymbol{n}}\underbrace{\sum_i \hat{\boldsymbol{e}}_i}_{=\boldsymbol{0}} = \bar{\boldsymbol{Y}}$$

3.12 Since $D_1 + D_2 = \mathbf{1}$, D_1, D_2, $\mathbf{1}$ are linearly dependent. So $X = [D_1, D_2, \mathbf{1}]$ is not of full rank, thus $X'X$ is not invertible. It follows that OLS ($(X'X)^{-1}X'Y$) is not well-defined for the first equation.

However, parameters in the second and third equations can be estimated by OLS.

(a) In terms of the best linear predictor, the second and third regression equations are the same. However, the coefficients are different.

$$\begin{aligned} y &= D_1\alpha_1 + D_2\alpha_2 + e = D_1\alpha_1 + (\mathbf{1} - D_1)\,\alpha_2 + e \\ &= \alpha_2 + D_1\,(\alpha_1 - \alpha_2) + e. \end{aligned}$$

Therefore, $\mu = \alpha_2$, $\phi = \alpha_1 - \alpha_2$.

(b) $\mathbf{1}'D_1 = $ (number of 1 in D_1) $= n_1$, $\mathbf{1}'D_2 = n_2$ similarly.

3.13 (a) Without loss of generality we can reorder the n observations such that the first n_1th observations are men, i.e., $d_1 = (1, 1, \ldots, 1, 0, 0, \ldots, 0)' = \left(\mathbf{1}_{n_1}', 0'\right)'$, $d_2 = \left(0', \mathbf{1}_{n_2}'\right)'$, and

$$Y = \left(Y_{n_1}', Y_{n_2}'\right)' = d_1\gamma_1 + d_2\gamma_2 + u = X\alpha + u,$$

where $X = [d_1\ d_2]$. Thus,

$$\hat{\alpha} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \left(X'X\right)^{-1} X'y = \begin{pmatrix} \mathbf{1}_{n_1}'\mathbf{1}_{n_1} & 0 \\ 0 & \mathbf{1}_{n_2}'\mathbf{1}_{n_2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}_{n_1}'Y_{n_1} \\ \mathbf{1}_{n_2}'Y_{n_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i \\ \frac{1}{n_2}\sum_{i=n_1+1}^{n} Y_i \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{pmatrix}.$$

(b) Let $D = [D_1\ D_2]$. Then, $Y^* = M_D y$, $X^* = M_D X$. In words, this transformation demeans dependent variable (Y) among the sample of men and women. In the same way, the transformation X to X^* implies, for each column vector (regressor x_k), demean each vector among the sample of men and women, respectively.

(c) $\tilde{\beta} = \left(X^{*'}X^*\right)^{-1} X^*Y^* = \left(X'M_DX\right)^{-1} X'M_DY$. Also, by the residual regression formula, $\hat{\beta} = \left(X'M_DX\right)^{-1} X'M_DY$. Thus $\hat{\beta} = \tilde{\beta}$. (This is known as Frisch-Waugh-Lovell Theorem).

3.16 A simple way to approach this problem is to say that by adding more regressors to the OLS problem, the sum of squared residuals (SSR) will be weakly smaller - this is a basic result from optimization theory. Therefore, given that $R^2 = 1 - SSR/SST$, where $SST = \sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^2$, by adding more regressors the SSR weakly decreases and SST is constant, so R^2 weakly increases. Equality holds when $\hat{\beta}_2 = 0$.

A more detailed solution would be: let $X = [X_1\ X_2]$, P_X, P_{X_1} are the projection matrices for X, X_1, and similarly for M_X, M_{X_1}. Then

$$R_1^2 = 1 - \frac{\tilde{e}'\tilde{e}}{\left(Y - 1\bar{Y}\right)'\left(Y - 1\bar{Y}\right)} = 1 - \frac{Y'M_{X_1}M_{X_1}Y}{\left(Y - 1\bar{Y}\right)'\left(Y - 1\bar{Y}\right)} = 1 - \frac{Y'M_{X_1}Y}{\left(Y - 1\bar{Y}\right)'\left(Y - 1\bar{Y}\right)}.$$

Similarly, $R_2^2 = 1 - \frac{Y'M_XY}{\left(Y-1\bar{Y}\right)'\left(Y-1\bar{Y}\right)}$.

$$\begin{aligned} R_2^2 - R_1^2 = \frac{Y'\left(M_{X_1} - M_X\right)Y}{\left(Y - 1\bar{Y}\right)'\left(Y - 1\bar{Y}\right)} &\geq 0 \\ \Leftrightarrow\ Y'\left(M_{X_1} - M_X\right)Y &\geq 0 \\ \Leftrightarrow\ Y'\left(P_X - P_{X_1}\right)Y &\geq 0. \end{aligned}$$

Note that $\left(P_X - P_{X_1}\right)\left(P_X - P_{X_1}\right) = P_XP_X - P_XP_{X_1} - P_{X_1}P_X + P_{X_1}P_{X_1}$. Note $P_XP_{X_1} = P_XX_1\left(X_1'X_1\right)^{-1}X_1' = X_1\left(X_1'X_1\right)^{-1}X_1' = P_{X_1}$ and $P_XP_X = P_X$ and $P_{X_1}P_{X_1} = P_{X_1}$. Thus $\left(P_X - P_{X_1}\right)$ is idempotent. You can also easily check $\left(P_X - P_{X_1}\right)$ is symmetric. So, $Y'\left(P_X - P_{X_1}\right)Y = [\left(P_X - P_{X_1}\right)Y]'[\left(P_X - P_{X_1}\right)Y] = \|\left(P_X - P_{X_1}\right)Y\|^2 \geq 0$ (symmetric idempotent matrix characteristic roots are either 0 or 1, so the quadratic form is positive semi-definite).

$R_2^2 = R_1^2$ when $\hat{\beta}_2 = 0$. To see this, let $\hat{\beta} = (X'X)^{-1}X'Y$ and $\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'Y$. From partitioned regression, recall that $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(Y - X_2\hat{\beta}_2) = (X_1'X_1)^{-1}X_1'Y = \tilde{\beta}_1$ when $\hat{\beta}_2 = 0$. Then, $\hat{Y} = X\hat{\beta} = X_1\tilde{\beta}_1 = \tilde{Y}$, and $R_2^2 = R_1^2$.

3.21 Let $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$ and $M_2 = I_n - X_2(X_2'X_2)^{-1}X_2'$. By the Frisch-Waugh-Lovell Theorem, the OLS estimators from the first equation can be expressed as: $\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2Y$ and $\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1Y$. Since M_1 and M_2 are symmetric idempotent matrices, we can define $\tilde{X}_1 = M_2X_1$ and $\tilde{X}_2 = M_1X_2$ and rewrite the least squares estimators as

$$\hat{\beta}_1 = \left(\tilde{X}_1'\tilde{X}_1\right)^{-1}\tilde{X}_1'Y \quad \text{and} \quad \hat{\beta}_2 = \left(\tilde{X}_2'\tilde{X}_2\right)^{-1}\tilde{X}_2'Y$$

Now consider the expressions for the "one regressor at a time" regressions:

$$\tilde{\beta}_1 = \left(X_1'X_1\right)^{-1}X_1'Y \quad \text{and} \quad \tilde{\beta}_2 = \left(X_2'X_2\right)^{-1}X_2'Y$$

So $\hat{\beta}_1 = \tilde{\beta}_1$ if $\tilde{X}_1 = X_1$ and $\hat{\beta}_2 = \tilde{\beta}_2$ if $\tilde{X}_2 = X_2$. Since $\tilde{X}_1 = \left[M_2 = I_n - X_2(X_2'X_2)^{-1}X_2'\right]X_1 = X_1 - X_2(X_2'X_2)^{-1}X_2'X_1$, we will have $\hat{\beta}_1 = \tilde{\beta}_1$ if $\underline{X_2'X_1 = 0}$ and this condition will also imply $\hat{\beta}_2 = \tilde{\beta}_2$.

3.22

$$\begin{aligned}
\tilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'Y \\
\tilde{u} &= Y - X_1\tilde{\beta}_1 \\
\tilde{\beta}_2 &= (X_2'X_2)^{-1}X_2'\tilde{u} \;=\; (X_2'X_2)^{-1}X_2'(Y - X_1\tilde{\beta}_1) \\
\hat{\beta}_2 &= (X_2'X_2)^{-1}X_2'(Y - X_1\hat{\beta}_1) \qquad \text{(from Partitioned Regression)}
\end{aligned}$$

So, $\tilde{\beta}_2 = \hat{\beta}_2$ if $\tilde{\beta}_1 = \hat{\beta}_1$ (which, in general, does not hold). $\tilde{\beta}_1 = \hat{\beta}_1$ when $M_2X_1 = X_1$ which holds when $X_2'X_1 = 0$. So, in general, $\tilde{\beta}_2 \neq \hat{\beta}_2$, but $\tilde{\beta}_2 = \hat{\beta}_2$ when $X_2'X_1 = 0$.

3.23 $\hat{\beta} = (X'X)^{-1}X'Y$ and $\tilde{\beta} = (Z'Z)^{-1}Z'Y$. Note that $M_{X_1}(X_2 - X_1) = M_{X_1}X_2$, so

$$\tilde{\beta}_2 = [(X_2 - X_1)'M_{X_1}(X_2 - X_1)]^{-1}(X_2 - X_1)'M_{X_1}Y = [X_2'M_{X_1}X_2]^{-1}X_2'M_{X_1}Y = \hat{\beta}_2.$$

And, by Partitioned Regression,

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'[Y - (X_2 - X_1)\tilde{\beta}_2] = \underbrace{(X_1'X_1)^{-1}X_1'(Y - X_2\hat{\beta}_2)}_{=\hat{\beta}_1} + (X_1'X_1)^{-1}X_1'X_1\hat{\beta}_2 = \hat{\beta}_1 + \hat{\beta}_2.$$

So,

$$\tilde{Y} = X_1\tilde{\beta}_1 + (X_2 - X_1)\tilde{\beta}_2 = X_1(\hat{\beta}_1 + \hat{\beta}_2) + (X_2 - X_1)\hat{\beta}_2 = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 = \hat{Y}$$

It follows that $\hat{\varepsilon} = \tilde{\varepsilon}$, and so $\hat{\sigma}^2 = \tilde{\sigma}^2$.

7 (a)

$$\begin{aligned}
\begin{pmatrix} E(\hat{\beta}_1|X) \\ E(\hat{\beta}_2|X) \end{pmatrix} &= E\left(\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}|X\right) = E\left(\hat{\beta}|X\right) = E\left((X'X)^{-1}X'y|X\right) \\
&= (X'X)^{-1}X'E\left(y|X\right) = (X'X)^{-1}X'X\beta = \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}
\end{aligned}$$

where the third equality uses the OLS "formula", the fourth equality uses the conditioning on X in the expectation, the fifth equality uses the assumption $E(y|X) = X\beta$. Looking at the first $k_1 \times 1$ subvector, we see $E(\hat{\beta}_1|X) = \beta_1$.

OR

$\hat{\beta}_1 = (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}y$, where $M_{X_2} = I - X_2(X_2'X_2)^{-1}X_2'$

$E(\hat{\beta}_1|X) = (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}E(y|X) = (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}X\beta$

$= (X_1'M_{X_2}X_1)^{-1}X_1'M_{X_2}(X_1\beta_1 + X_2\beta_2) = \beta_1 + 0 = \beta_1$

(b) $\hat{\hat{\beta}}_1 = (X_1'X_1)^{-1}X_1'\hat{y} = (X_1'X_1)^{-1}X_1'X\hat{\beta}$. From part (a), $E(\hat{\hat{\beta}}_1|X) = E\left((X_1'X_1)^{-1}X_1'X\hat{\beta}|X\right) = (X_1'X_1)^{-1}X_1'XE\left(\hat{\beta}|X\right) = (X_1'X_1)^{-1}X_1'X\beta = (X_1'X_1)^{-1}X_1'(X_1 \;\; X_2)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$
$= (X_1'X_1)^{-1}X_1'\left(X_1\beta_1 + X_2\beta_2\right) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2$.

In general, $\hat{\hat{\beta}}_1$ is not unbiased, unless $\beta_2 = 0$ (or $X_1'X_2 = 0$ almost surely).

(c) Note that

$$\begin{pmatrix} I_{k_1} & 0\\ 0 & I_{k_2}\end{pmatrix} = I_k = (X'X)^{-1}X'X = (X'X)^{-1}X'[X_1 \;\; X_2] = [(X'X)^{-1}X'X_1 \;\; (X'X)^{-1}X'X_2]$$

So, $(X'X)^{-1}X'X_1 = \begin{pmatrix} I_{k_1}\\ 0\end{pmatrix}$.
$\tilde{\tilde{\beta}} = (X'X)^{-1}X'\tilde{y} = (X'X)^{-1}X'X_1\tilde{\beta}_1 = \begin{pmatrix} I_{k_1}\\ 0\end{pmatrix}\tilde{\beta}_1 = \begin{pmatrix} \tilde{\beta}_1\\ 0\end{pmatrix}$

(d) Form the residuals, $\hat{\varepsilon} = \tilde{y} - X\tilde{\tilde{\beta}} = (X_1\tilde{\beta}_1) - [X_1 \;\; X_2]\begin{pmatrix} \tilde{\beta}_1\\ 0\end{pmatrix} = (X_1\tilde{\beta}_1) - (X_1\tilde{\beta}_1) = 0$.
So, $\hat{\varepsilon}'\hat{\varepsilon} = 0$, and $R^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(\tilde{y}-\bar{\tilde{y}}\ell)'(\tilde{y}-\bar{\tilde{y}}\ell)} = 1$.

(e) First, note that $\text{Var}(\tilde{\beta}_1|X) = \text{Var}((X_1'X_1)^{-1}X_1'y|X) = (X_1'X_1)^{-1}X_1'\text{Var}(y|X)X_1(X_1'X_1)^{-1} = (X_1'X_1)^{-1}X_1'\sigma^2 I X_1(X_1'X_1)^{-1} = \sigma^2(X_1'X_1)^{-1}X_1'X_1(X_1'X_1)^{-1} = \sigma^2(X_1'X_1)^{-1}$
$\text{Var}(\tilde{\tilde{\beta}}|X) = \text{Var}\left(\begin{pmatrix} I_{k_1}\\ 0\end{pmatrix}\tilde{\beta}_1|X\right) = \begin{pmatrix} I_{k_1}\\ 0\end{pmatrix}\text{Var}(\tilde{\beta}_1|X)\begin{pmatrix} I_{k_1}\\ 0\end{pmatrix}' = \begin{pmatrix} \text{Var}(\tilde{\beta}_1|X) & 0\\ 0 & 0\end{pmatrix}$
$= \begin{pmatrix} \sigma^2(X_1'X_1)^{-1} & 0\\ 0 & 0\end{pmatrix}$.