ECON 709B - Problem Set 2

Alex von Hafften[*]

11/22/2020

1. 3.2[1] Consider the OLS regression of the $n \times 1$ vector y on the $n \times k$ matrix X. Consider an alternative set of regressors $Z = XC$, where C is a $k \times k$ non-singular matrix. Thus, each column of Z is a mixture of some of the columns of X. Compare the OLS estimates and residuals from the regression of Y on X to the OLS estimates from the regression of y on Z.

The OLS estimates and residuals from the regression of y on X:

$$\hat{\beta}_X = (X'X)^{-1}X'y$$

$$\hat{e}_X = Me = (I - X(X'X)^{-1}X')e$$

The OLS estimates and residuals from the regression of y on Z:

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

$$
\begin{aligned}
\hat{e}_Z &= M_Z e \\
&= (I - Z(Z'Z)^{-1}Z')e \\
&= (I - (XC)((XC)'(XC))^{-1}(XC)')e \\
&= (I - (XC)(C'X'XC)^{-1}C'X')e \\
&= (I - (XC)(C^{-1})(X'X)^{-1}(C')^{-1}C'X')e \\
&= (I - X(X'X)^{-1}X')e
\end{aligned}
$$

Thus, the OLS estimates from the regression of y on Z are those from the regression of y on X pre-multiplied by C^{-1} and the residuals are the same in both regressions.

---

2. 3.5 Let $\hat{e}$ be the OLS residual from a regression of y on $X = [X_1 X_2]$. Find $X_2'\hat{e}$.

Note that $X_2 = X\Gamma_2$ where Γ_2 is the last k_2 columns of a I_k, so it is $k \times k_2$:

$$X_2'\hat{e} = (X\Gamma_2)'\hat{e} = \Gamma_2'X'\hat{e} = \Gamma_2'0 = 0$$

3.6 Let $\hat{y} = X(X'X)^{-1}X'y$. Find the OLS coefficient from a regression of $\hat{y}$ on X.

Let $\hat{\beta} = (X'X)^{-1}X'y$ be the OLS coefficient from a regression of y on X. Thus, the OLS coefficient from a regression of $\hat{y}$ on X is

$$
\begin{aligned}
\tilde{\beta} &= (X'X)^{-1}X'\hat{y} \\
&= (X'X)^{-1}X'X(X'X)^{-1}X'y \\
&= (X'X)^{-1}X'y \\
&= \hat{\beta}
\end{aligned}
$$

3.7 Show that if $X = [X_1 \ X_2]$, then $PX_1 = X_1$ and $MX_1 = 0$.

Note that $X_1 = X\Gamma_1$ where Γ_1 is the first k_1 columns of a I_k, so it is $k \times k_1$:

$$PX_1 = PX\Gamma_1 = X(X'X)^{-1}X'X\Gamma_1 = X\Gamma_1 = X_1$$

$$MX_1 = (I_n - P)X_1 = I_nX_1 - PX_1 = X_1 - X_1 = 0$$

3. 3.11 Show that when X contains a constant $\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y}$.

$$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{e}_i) = \frac{1}{n}\sum_{i=1}^{n}y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i = \bar{y} - \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i$$

We know from exercise 3.5 that $X_1'\hat{e} = 0$ where $X = [X_1 \ X_2]$. Choose X_1 be the column of ones representing the constant, so $\sum_{i=1}^{n}\hat{e}_i = 0 \implies \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y}$.

3.12 A dummy variable takes on only the values 0 and 1. It is used for categorical data, such as an individual's gender. Let D_1 and D_2 be vectors of 1's and 0's, with the ith element of D_1 equaling 1 and that of D_2 equaling 0 if the person is a man, and the reserve if the person is a woman. Suppose that there are n_1 men and n_2 women in the sample. Consider fitting the following three equations by OLS: (3.53) $y = \mu + D_1\alpha_1 + D_2\alpha_2 + e$, (3.54) $y = D_1\alpha_1 + D_2\alpha_2 + e$, and (3.55) $y = \mu + D_1\phi + e$. Can all three equations be estimated by OLS? Explain if not.

If gender is binary and all people in the sample identify either as a man or woman, then only (3.54) and (3.55) can be estimated using OLS. In (3.53) X does not have full ($\text{rank}(X) = 1 \neq 2$) because $D_1 = 1_n - D_2$, so $X'X$ is not invertible.

If gender is not binary, so $D_1 \neq 1_n - D_2$, then all three equations can be estimated using OLS.

(a) Compare regressions (3.54) and (3.55). Is one more general than the other? Explain the relationship between the parameters in (3.54) and (3.55).

(3.54) and (3.55) result in estimates that related and the same residuals, but (3.55) is more general than (3.54) because it includes a constant, so if more variables are added it ensures that the regression line passes through the sample averages and that R^2 have a helpful interpretation.

α_1 is the average of y for men and α_2 is the average of y for women.

μ is the average of y for women and ϕ is the difference between the average y for men and women.

So $\mu = \alpha_2$ and $\phi = \alpha_1 - \mu = \alpha_1 - \alpha_2$.

(b) Compute $1'_n D_1$ and $1'_n D_2$, where 1_n is a $n \times 1$ vector of ones.

$$1'_n D_1 = n_1$$
$$1'_n D_2 = n_2$$

3.13 Let D_1 and D_2 be defined as in the previous exercise.

(a) In the OLS regression $Y = D_1 \hat{\gamma}_1 + D_2 \hat{\gamma}_2 + \hat{u}$. Show that $\hat{\gamma}_1$ is the sample mean of the dependent variance among the men in the sample ($\bar{y}_1$) and that $\hat{\gamma}_2$ is the sample mean the women in the sample ($\bar{y}_2$)

$$
\begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \left(\begin{pmatrix} D'_1 \\ D'_2 \end{pmatrix} (D_1 \ D_2) \right)^{-1} \begin{pmatrix} D'_1 \\ D'_2 \end{pmatrix} y
$$
$$
= \begin{pmatrix} D'_1 D_1 & D'_1 D_2 \\ D'_2 D_1 & D'_2 D_2 \end{pmatrix}^{-1} \begin{pmatrix} D'_1 y \\ D'_2 y \end{pmatrix}
$$
$$
= \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}^{-1} \begin{pmatrix} D'_1 y \\ D'_2 y \end{pmatrix}
$$
$$
= \begin{pmatrix} (D'_1 y)/n_1 \\ (D'_2 y)/n_2 \end{pmatrix}
$$
$$
= \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}
$$

(b) Let X $(n \times k)$ be an additional matrix of regressors. Describe in words the transformations $y^* = y - D_1 \bar{y}_1 - D_2 \bar{y}_2$ and $X^* = X - D_1 \bar{X}'_1 - D_2 \bar{X}'_2$ where $\bar{X}'_1$ and $\bar{X}'_1$ are the $k \times 1$ means of the regressors for men and women, respectively.

For men ($D_1 = 1$ and $D_2 = 0$), the transformations become:

$$y^* = y - D_1 \bar{y}_1$$
$$X^* = X - D_1 \bar{X}'_1$$

For women ($D_1 = 0$ and $D_2 = 1$), the transformations become:

$$y^* = y - D_2 \bar{y}_2$$
$$X^* = X - D_2 \bar{X}'_2$$

Thus, these transformations demean the dependent variable and the regressors using the group means.

4. 3.16 Consider two least squares regressions $y = X_1 \tilde{\beta}_1 + \tilde{e}$ and $y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{e}$. Let R_1^2 and R_2^2 be the R-squared from the two regressions. Show that $R_2^2 \geq R_1^2$. Is there a case (explain) when these is equality $R_2^2 = R_1^2$?

Notice that both R_1^2 and R_1^2 have the same denominator $\sum_{i=1}^{n}(y_i - \bar{y})^2$. Thus, we can show that $R_1^2 \leq R_2^2$ by showing $\sum_{i=1}^{n} \tilde{e}^2 \geq \sum_{i=1}^{n} \hat{e}^2$.

By the Frisch-Waugh-Lovell Theorem, $\hat{e}$ may be computed by regressing y on X_1 to obtain residuals $\tilde{e}$, regressing X_2 on X_1 to obtain residuals $\hat{X}_2$, and regressing $\tilde{e}$ on $\hat{X}_2$ to obtaining fitted values $\tilde{Z}$ and residuals $\hat{e}$. Thus,

$$\tilde{e}'\tilde{e} = (\tilde{Z} + \hat{e})'(\tilde{Z} + \hat{e}) = \tilde{Z}'\tilde{Z} + 2\tilde{Z}'\hat{e} + \hat{e}'\hat{e} = \tilde{Z}'\tilde{Z} + \hat{e}'\hat{e} \geq \hat{e}'\hat{e}$$

Because $\tilde{Z}$ and $\hat{e}$ are orthogonal.

Yes, $R_2^2 = R_1^2$ if X_2 does not contain any information that helps explain variation in y. That is, X_2 and y are orthogonal.

5. 3.21 Consider the least squares regression estimators $y_i = X_{1i}\hat{\beta}_1 + X_{2i}\hat{\beta}_2 + \hat{e}_i$ and the "one regressor at a time" regression estimators $y_i = X_{1i}\tilde{\beta}_1 + \tilde{e}_{1i}$ and $y_i = X_{2i}\tilde{\beta}_2 + \tilde{e}_{2i}$. Under what condition does $\tilde{\beta}_1 = \hat{\beta}_1$ and $\tilde{\beta}_2 = \hat{\beta}_2$?

If X_1 and X_2 are orthogonal, $\tilde{\beta}_1 = \hat{\beta}_1$ and $\tilde{\beta}_2 = \hat{\beta}_2$. By the Frisch-Waugh-Lovell Theorem, $\hat{\beta}_2$ may be computed by regressing y on X_1 to obtaining $\tilde{\beta}_1$ residuals $\tilde{e}$, regressing X_2 on X_1 to obtain residuals $\hat{X}_2$, and regressing $\tilde{e}$ on $\hat{X}_2$ to obtaining $\hat{\beta}_2$. If X_1 and X_2 are orthogonal, then $\tilde{X}_2 = X_2$, so $\tilde{\beta}_1 = \hat{\beta}_1$. Similar logic applies for $\tilde{\beta}_2 = \hat{\beta}_2$.

3.22 You estimate a least squares regression $y_i = X_{1i}'\tilde{\beta}_1 + \tilde{u}_i$ and then regress the residuals on another set of regressors $\tilde{u}_i = X_{2i}'\tilde{\beta}_2 + \tilde{e}_i$. Does this second regression give you the same estimated coefficients as from estimation of a least squares regression on both set of regressors? $y_i = X_{1i}'\hat{\beta}_1 + X_{2i}'\hat{\beta}_2 + \hat{e}_i$. In other words, is it true that $\tilde{\beta}_2 = \hat{\beta}_2$? Explain your reasoning.

Not necessarily. Only if X_1 and X_2 are orthogonal, $\tilde{\beta}_2 = \hat{\beta}_2$. We need to consider the comovement between X_1 and X_2. If there's no comovement (they are orthogonal), then we can regress them one at a time.

3.23 The data matrix is (y, X) with $X = [X_1 \ X_2]$, and consider the transformed regressor matrix $Z = [X_1, X_2 - X_1]$. Suppose you do a least squares regression of y on X, and a least squares regression of y on Z. Let $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ denote the residual variance estimates from the two regressions. Give a formula relating $\hat{\sigma}^2$ and $\tilde{\sigma}^2$? (Explain your reasoning.)

We show that $\hat{\sigma}^2 = \tilde{\sigma}^2$ by showing that the projection matrix for the regression of y on X and the projection matrix for the regression of y on Z are equal thus the annilihator matrices are equal and thus the residuals are equal.

The projection matrix for the regression of y on X is

$$P_X = X(X'X)^{-1}X'$$
$$= [X_1 \ X_2]([X_1 \ X_2]'[X_1 \ X_2])^{-1}[X_1 \ X_2]'$$
$$= [X_1 \ X_2] \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} [X_1 \ X_2]'$$
$$= [X_1 \ X_2] \begin{bmatrix} \frac{X_2'X_2}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} & -\frac{X_1'X_2}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} \\ -\frac{X_2'X_1}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} & \frac{X_1'X_1}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} \end{bmatrix} [X_1 \ X_2]'$$
$$= \begin{bmatrix} \frac{X_1X_2'X_2 - X_2X_2'X_1}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} & \frac{X_2X_1'X_1 - X_1X_1'X_2}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} \end{bmatrix} [X_1 \ X_2]'$$
$$= \frac{X_1X_2'X_2X_1' - X_2X_2'X_1X_1' + X_2X_1'X_1X_2' - X_1X_1'X_2X_2'}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1}$$

Thus, the projection matrix for the regression of y on Z is

$$P_Z = \frac{Z_1Z_2'Z_2Z_1' - Z_2Z_2'Z_1Z_1' + Z_2Z_1'Z_1Z_2' - Z_1Z_1'Z_2Z_2'}{Z_1'Z_1Z_2'Z_2 - Z_1'Z_2Z_2'Z_1}$$

The numerator of P_Z simplifies to the numerator of P_X:

$$Z_1Z_2'Z_2Z_1' - Z_2Z_2'Z_1Z_1' + Z_2Z_1'Z_1Z_2' - Z_1Z_1'Z_2Z_2'$$
$$= X_1(X_2 - X_1)'(X_2 - X_1)X_1' - (X_2 - X_1)(X_2 - X_1)'X_1X_1' + (X_2 - X_1)X_1'X_1(X_2 - X_1)'$$
$$- X_1X_1'(X_2 - X_1)(X_2 - X_1)'$$
$$= X_1(X_2'X_2 - 2X_1'X_2 + X_1'X_1)X_1' - (X_2X_2' - X_1X_2' - X_2X_1' + X_1X_1')X_1X_1'$$
$$+ (X_2X_1'X_1 - X_1X_1'X_1)(X_2 - X_1)' - X_1X_1'(X_2X_2' - X_1X_2' - X_2X_1' + X_1X_1')$$
$$= X_1X_2'X_2X_1' - 2X_1X_1'X_2X_1' + X_1X_1'X_1X_1' - X_2X_2'X_1X_1' + X_1X_2'X_1X_1' + X_2X_1'X_1X_1' - X_1X_1'X_1X_1'$$
$$+ X_2X_1'X_1X_2' - X_1X_1'X_1X_2' - X_2X_1'X_1X_1' + X_1X_1'X_1X_1' - X_1X_1'X_2X_2' + X_1X_1'X_1X_2' + X_1X_1'X_2X_1' - X_1X_1'X_1X_1'$$
$$= X_1X_2'X_2X_1' - X_2X_2'X_1X_1' + X_2X_1'X_1X_2' - X_1X_1'X_2X_2'$$

The denominator of P_Z simplifies to the denominator of P_X:

$$Z_1'Z_1Z_2'Z_2 - Z_1'Z_2Z_2'Z_1$$
$$= X_1'X_1(X_2 - X_1)'(X_2 - X_1) - X_1'(X_2 - X_1)(X_2 - X_1)'X_1$$
$$= X_1'X_1(X_2'X_2 - 2X_1'X_2 + X_1'X_1) - X_1'(X_2X_2' - X_1X_2' - X_2X_1' + X_1X_1')X_1$$
$$= (X_1'X_1X_2'X_2 - 2X_1'X_1X_1'X_2 + X_1'X_1X_1'X_1) - (X_1'X_2X_2'X_1 - X_1'X_1X_2'X_1 - X_1'X_2X_1'X_1 + X_1'X_1X_1'X_1)$$
$$= X_1'X_1X_2'X_2 - 2X_1'X_1X_1'X_2 + X_1'X_1X_1'X_1 - X_1'X_2X_2'X_1 + X_1'X_1X_2'X_1 + X_1'X_2X_1'X_1 - X_1'X_1X_1'X_1$$
$$= X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1$$

Thus, P_Z and P_X are equal:

$$P_Z = \frac{X_1X_2'X_2X_1' - X_2X_2'X_1X_1' + X_2X_1'X_1X_2' - X_1X_1'X_2X_2'}{X_1'X_1X_2'X_2 - X_1'X_2X_2'X_1} = P_X$$

Thus, $M_Z = I - P_Z = I - P_X = M_X \implies \tilde{e} = M_Zy = M_Xy = \hat{e} \implies \tilde{\sigma}^2 = \hat{\sigma}^2$.

6. 3.24 Use the data set from Section 3.22 and the sub-sample used for equation (3.50) (see Section 3.25 for data construction).

(a) Estimate equation (3.50) and compute the equation R^2 and sum of squared errors.

```r
library(tidyverse)
```

```r
cps09mar <- read_delim("cps09mar.txt",
                       delim = "\t",
                       col_names = c("age", "female", "hisp", "education", "earnings",
                                     "hours", "week", "union", "uncov", "region", "race",
                                     "maritial"),
                       col_types = "dddddddddddd") %>%
  mutate(experience = age - education - 6,
         experience_2 = (experience^2)/100,
         wage = earnings / (hours*week),
         l_wage = log(wage),
         constant = 1) %>%
  filter(race == 4,
         maritial == 7,
         female == 0,
         experience < 45)
y <- cps09mar$l_wage
x <- cps09mar %>%
  select(education, experience, experience_2, constant) %>%
  as.matrix() %>%
  unname()
n <- dim(x)[1]
i <- diag(nrow = n, ncol = n)
# Estimate Equation 3.50
beta <- solve(t(x) %*% x) %*% t(x) %*% y
print(beta)
```

```
##              [,1]
## [1,]  0.14430729
## [2,]  0.04263326
## [3,] -0.09505636
## [4,]  0.53089068
```

```r
p_x <- x %*% solve(t(x) %*% x) %*% t(x)
m_x <- i - p_x
e_hat <- m_x %*% y
sse_a <- sum(e_hat^2)
print(sse_a)
```

```
## [1] 82.505
```

```r
y_bar <- mean(y)
r_squared_a <- 1 - sse_a/sum((y-y_bar)^2)
print(r_squared_a)
```

```
## [1] 0.3893207
```

(b) Re-estimate the slope on education using the residual regression approach. Regress log(wage) on experience and its square, regress education on experience and its square, and the residuals on the residuals. Report the estimates from this final regression, along with the equation R^2 and sum of squared errors. Does the slope coefficient equal the value in (3.50)? Explain.

```
x_1 <- cps09mar %>%
  select(experience, experience_2, constant) %>%
  as.matrix() %>%
  unname()
m_x_1 <- i - x_1 %*% solve(t(x_1) %*% x_1) %*% t(x_1)
z <- cps09mar$education
u_hat <- m_x_1 %*% z
v_hat <- m_x_1 %*% y
beta_1 <- (t(u_hat) %*% u_hat)^(-1) %*% t(u_hat) %*% v_hat
print(beta_1)
```

```
##            [,1]
## [1,] 0.1443073
```

```
m_u_hat <- i - u_hat %*% (t(u_hat) %*% u_hat)^(-1) %*% t(u_hat)
d_hat <- m_u_hat %*% v_hat
sse_b <- sum(d_hat^2)
print(sse_b)
```

```
## [1] 82.505
```

```
v_hat_bar <- mean(v_hat)
r_squared_b <- 1 - sse_b/sum((v_hat - v_hat_bar)^2)
print(r_squared_b)
```

```
## [1] 0.368738
```

(c) Are the R^2 and sum of squared errors from parts (a) and (b) equal? Explain.

The SSE is the same between (a) and (b) because the residuals are the same no matter if you do a partition regression or a regression with all variables at once. The R^2 in part (b) is lower than in part (a). Since the SSE is the same, both parts have the same denominator. Because there is some explanatory power from the first regression, the denominator is larger in part (a) than in part (b).

3.25 Estimate equation (3.50) as in part (a) of the previous question. Let $\hat{e}_i$ be the OLS residual, $\hat{Y}_i$ the predicted value from the regression, X_{1i} be education and X_{2i} be experience.

Numerically calculate the following:

```
x1 <- x[,1]
x2 <- x[,2]
p_x <- x %*% solve(t(x) %*% x) %*% t(x)
m_x <- i - p_x
y_hat <- p_x %*% y
e_hat <- m_x %*% y
```

(a) $\sum_{i=1}^{n} \hat{e}_i$

```
sum(e_hat)
```

[1] 1.479396e-12

(b) $\sum_{i=1}^{n} X_{1i}\hat{e}_i$

```
sum(x1 * e_hat)
```

[1] 2.403766e-11

(c) $\sum_{i=1}^{n} X_{2i}\hat{e}_i$

```
sum(x2 * e_hat)
```

[1] 4.451689e-12

(d) $\sum_{i=1}^{n} X_{1i}^2\hat{e}_i$

```
sum(x1^2 * e_hat)
```

[1] 133.1331

(e) $\sum_{i=1}^{n} X_{2i}^2\hat{e}_i$

```
sum(x2^2 * e_hat)
```

[1] -8.045165e-11

(f) $\sum_{i=1}^{n} \hat{Y}_i\hat{e}_i$

```
sum(y_hat * e_hat)
```

[1] 4.701786e-12

(g) $\sum_{i=1}^{n} X_{1i}\hat{e}_i^2$

```
sum(x1 * e_hat^2)
```

[1] 1240.049

Are these calculations consistent with the theoretical properties of OLS? Explain.

Yes. When working with a computer, there is always some rounding, so we should interpret very small numbers as zero in terms of consistency with the theoretical predictions.

(a) The sum of residuals is zero.
(b) Residuals are orthogonal to any partition of X.
(c) Residuals are orthogonal to any partition of X.
(d) [No particular theoretical property.]
(e) Note that squared experience is a column in X, so it is a partition of X. Residuals are orthogonal to any partition of X.

(f) Residuals are orthogonal to fitted values.
(g) [No particular theoretical property.]

7. Given the $n \times 1$ vector y and the $n \times k$ matrix X. Assume: $\text{rank}(X) = k$; $E(y|X) = X\beta$; and $\text{var}(y|X) = \sigma^2 I$. Partition X: $X = [X_1\ X_2]$ where X_1 is $n \times k_1$, X_2 is $n \times k_2$, and $k_1 + k_2 = k$. And similarly partition β: $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, where β_1 is $k_1 \times 1$ and β_2 is $k_2 \times 1$.

(a) Consider the OLS regression of y on X that yields the OLS estimator $\hat{\beta}$. What is $E[\hat{\beta}_1|X]$? Simplify your answer.

$$E[\hat{\beta}|X] = E[(X'X)^{-1}X'y|X] = (X'X)^{-1}X'E[y|X] = (X'X)^{-1}X'X\beta = \beta$$

Thus, $E[\hat{\beta}_1|X] = \beta_1$.

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

$$\begin{aligned} E[\hat{\hat{\beta}}_1|X] &= E[(X_1'X_1)^{-1}X_1'\hat{y}|X] \\ &= (X_1'X_1)^{-1}X_1'E[X\hat{\beta}|X] \\ &= (X_1'X_1)^{-1}X_1'X E[\hat{\beta}|X] \\ &= (X_1'X_1)^{-1}X_1'X\beta \\ &= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2) \\ &= (X_1'X_1)^{-1}X_1'X_1\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \\ &= \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 \end{aligned}$$

$\hat{\hat{\beta}}_1$ is generally not a unbiased estimator of β_1. If X_1 and X_2 are orthogonal ($X_1'X_2 = 0$) or $\beta_2 = 0$, then it is an unbiased estimator of β_1.

(c) Consider the OLS regression of y on X_1 that yields the OLS estimator $\tilde{\beta}_1$. Let $\tilde{y} = X_1\tilde{\beta}_1$. Now consider the OLS regression of $\tilde{y}$ on X that yields the OLS estimator $\tilde{\tilde{\beta}}$. How is $\tilde{\tilde{\beta}}$ related to $\tilde{\beta}_1$? (Provide a mapping between $\tilde{\tilde{\beta}}$ and $\tilde{\beta}_1$ that does not involve X.)

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$

$$\tilde{\tilde{\beta}} = (X'X)^{-1}X'\tilde{y} = (X'X)^{-1}X'X_1\tilde{\beta}_1 = \Gamma\tilde{\beta}_1$$

Notice that Γ is $k \times k_1$. It is a $k_1 \times k_1$ identity matrix with k_2 rows of zeros below:

$$\Gamma = \begin{pmatrix} I_{k_1} \\ 0_{k_2 \times k_1} \end{pmatrix}$$

(d) What is the R^2 for the OLS regression of $\tilde{y}$ on X (from part (c))? Simplify your answer.

The annihilator matrix is

$$M = I_n - X(X'X)X'$$

Thus, the residuals are

$$\begin{aligned}
\tilde{\tilde{e}} &= M\tilde{y} \\
&= (I_n - X(X'X)X')(X_1\tilde{\beta}_1) \\
&= X_1\tilde{\beta}_1 - X(X'X)X'X_1\tilde{\beta}_1 \\
&= X_1\tilde{\beta}_1 - X\Gamma\tilde{\beta}_1 \\
&= X_1\tilde{\beta}_1 - X_1\tilde{\beta}_1 \\
&= 0
\end{aligned}$$

Thus, $R^2 = 1$.

(e) What is $\text{var}(\tilde{\tilde{\beta}}|X)$? Simply your answer.

$$\begin{aligned}
\text{var}(\tilde{\beta}_1|X) &= \text{var}((X_1'X_1)^{-1}X_1'y|X) \\
&= (X_1'X_1)^{-1}X_1'\text{var}(y|X)((X_1'X_1)^{-1}X_1')' \\
&= (X_1'X_1)^{-1}X_1'\sigma^2 I((X_1'X_1)^{-1}X_1')' \\
&= \sigma^2(X_1'X_1)^{-1}X_1'X_1(X_1'X_1)^{-1} \\
&= \sigma^2(X_1'X_1)^{-1}
\end{aligned}$$

$$\begin{aligned}
\text{var}(\tilde{\tilde{\beta}}|X) &= \text{var}(\Gamma\tilde{\beta}_1|X) \\
&= \Gamma\text{var}(\tilde{\beta}_1|X)\Gamma' \\
&= \Gamma\sigma^2(X_1'X_1)^{-1}\Gamma' \\
&= \sigma^2\Gamma(X_1'X_1)^{-1}\Gamma' \\
&= \sigma^2 \begin{pmatrix} (X_1'X_1)^{-1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & 0_{k_2 \times k_2} \end{pmatrix}
\end{aligned}$$