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-multipled by C^{-1} and the residuals are the same in both regressions.

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

[1]These problems come from *Econometrics* by Bruce Hansen, revised on October 23, 2020.

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 $k \times k_1$:

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

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

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 indentify 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 (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 excerise.

(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 ($\bar{y}_2$)

...

(b)

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 $R_2^2 = R_1^2$?

...

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 $\tilde{\beta}_2 = \hat{\beta}_2$?

...

3.22

...

3.23

...

6. 3.24

...

3.25

...

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.

From lecture, we have that

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2)$$

$$\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1y$$

So,

$$E[\hat{\beta}_2|X] = E[(X_2'M_1X_2)^{-1}X_2'M_1y|X] =$$

and

$$E[\hat{\beta}_1|X] = E[(X_1'X_1)^{-1}X_1'(y - X_2\hat{\beta}_2)|X] = E[(X_1'X_1)^{-1}X_1'y|X] - E[(X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2|X] = (X_1'X_1)^{-1}X_1'E[y|X] - (X_1'X_1)^{-1}$$

$$\hat{\beta} = (X'X)^{-1}X'y \implies \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} [X_1 \; X_2] \right)^{-1} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} y = \left(\begin{pmatrix} X_1X_1 & X_1 \\ X_2X_1 & \end{pmatrix} [X_1 \; X_2] \right)^{-1} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} y$$

(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 $E[\hat{\hat{\beta}}_1|X]$? (Simpllify your answer.) Is $\hat{\hat{\beta}}_1$ 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.)

...

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

...

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