Econometrics 710 - Problem Set # 3

February 18, 2015

TA: Cristian Hernandez
cahernandez3@wisc.edu

Exercise 3.1

$$0 = \bar{g}_n(\hat{\mu}, \hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n} g(y_i, \hat{\mu}, \hat{\sigma}^2) = \begin{pmatrix} \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{\mu}) \\ \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{\mu})^2 - \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \bar{y} - \hat{\mu} \\ \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{\mu})^2 - \hat{\sigma}^2 \end{pmatrix}$$

Thus, $\hat{\mu} = \bar{y}$, and $\hat{\sigma^2} = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \bar{y})^2$.

Exercise 3.2

All column vectors of $\mathbf{Z}$ are a linear combination of the column vectors of $\mathbf{X}$. Thus the column space of $\mathbf{Z}$ is the same as the column space of $\mathbf{X}$ with the same dimension k. Thus fitted values and residuals from regression of $\mathbf{y}$ on $\mathbf{X}$ are the same as those from the regression of $\mathbf{y}$ on $\mathbf{Z}$.

The OLS estimate from the regression of $\mathbf{y}$ on $\mathbf{X}$ is $\hat{\beta} = (X'X)^{-1}X'y$ and the fitted values from the regression are $X\hat{\beta}$. Residual vector is $\hat{e} = My$, where $M = I - X(X'X)^{-1}X'$ denotes orthogonal projection matrix. The OLS estimates from the regression of $\mathbf{y}$ on $\mathbf{Z}$ is

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

Therefore, fitted values $Z\tilde{\beta}$ are equal to $X\hat{\beta}$ since $Z\tilde{\beta} = XCC^{-1}\hat{\beta} = X\hat{\beta}$. Thus residuals are equal. (You can also check $M_z = I_n - Z(Z'Z)^{-1}Z' = M$ by matrix calculation, so that $\hat{e} = My = M_z y = \tilde{e}$.)

Exercise 3.3

Note that $MX = (I - X(X'X)^{-1}X')X = 0$

$$X'\hat{e} = X'My = (MX)'y = \mathbf{0}_{k\times 1}$$

Exercise 3.4

Let X_1, X_2 are $n \times k_1, n \times k_2$ matrix where $k_1 + k_2 = k$

$X'\hat{e} = \begin{pmatrix} X_1'\hat{e} \\ X_2'\hat{e} \end{pmatrix} = \mathbf{0}_{k\times 1}$ by exercise 3.3. Threfore, $X_2'\hat{e} = \mathbf{0}_{k_2\times 1}$.

Exercise 3.5

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

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

$X'\hat{e} = 0$ by exercise 3.3.

Exercise 3.6

Note that $PX = X$. The OLS coefficient from a regression of $\hat{y}$ on X is,

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

(or using exercise 3.5, $(X'X)^{-1}X'(y - \hat{e}) = (X'X)^{-1}X'y$)

Exercise 3.7

Note that $PX = X, MX = 0$. By working out the partitioned matrix, $[X_1 \quad X_2] = X = PX = [PX_1 \quad PX_2], MX = [MX_1 \quad MX_2]$. Therefore $PX_1 = X_1$, and $MX_1 = 0$

Exercise 3.8

$MM = (I - P)(I - P) = I - 2P + P^2 = I - P = M$ since P is idempotent, $P^2 = PP$

Exercise 3.9

$tr(M) = tr(I_n - P) = tr(I_n) - tr(P) = n - tr(X(X'X)^{-1}X') = n - tr((X'X)^{-1}X'X) = n - tr(I_k) = n - k$

Here we use the properties of trace; $tr(A + B) = tr(A) + tr(B), tr(AB) = tr(BA)$, and $tr(I_k) = k$

Exercise 3.10

$$\begin{aligned}
P = X(X'X)^{-1}X' &= \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \\
&= \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{pmatrix}^{-1} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \quad (\text{since } X_1'X_2 = 0) \\
&= \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2'X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \\
&= X_1(X_1'X_1)^{-1}X_1' + X_2(X_2'X_2)^{-1}X_2' = P_1 + P_2
\end{aligned}$$

Exercise 3.11

When X contains a constant, $\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i = 0$ by exercise 3.3. Therefore,

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

Exercise 3.12

Since $\mathbf{d}_1 + \mathbf{d}_2 = \iota$, $\mathbf{d}_1, \mathbf{d}_2, \iota$ are linearly dependent. So $\mathbf{X} = [\mathbf{d}_1, \mathbf{d}_2, \iota]$ are not of full rank, thus not invertible. Thus, we cannot identify all three parameters in (3.46) by OLS estimates, which is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. However, parameters in (3.47), (3.48) can be estimated by OLS.

(a) In terms of best linear predictor, regressions (3.47) and (3.48) are same. However, the coefficients of the regressions are different, so we need different interpretations.

To see this, note that $\mathbf{d}_1 + \mathbf{d}_2 = \iota$,

$$\begin{aligned} y &= \mathbf{d}_1\alpha_1 + \mathbf{d}_2\alpha_2 + \mathbf{e} = \mathbf{d}_1\alpha_1 + (\iota - \mathbf{d}_1)\alpha_2 + \mathbf{e} \\ &= \alpha_2 + \mathbf{d}_1(\alpha_1 - \alpha_2) + \mathbf{e} \end{aligned}$$

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

(b) $\iota' d_1 = (\#\text{ of 1 in } \mathbf{d}_1) = n_1, \iota' d_2 = n_2$

(c) Definitely the assumption $\mathbb{E}(\mathbf{x}_i e_i) = \mathbf{0}$ holds true. Since α is the best linear predictor and all the regressors are linearly independent, by construction $\mathbb{E}(\mathbf{x}_i e_i) = \mathbf{0}$.

Exercise 3.13

(a) Without loss of generality we can reorder n observations in the way that first n_1th observation are men, i.e $\mathbf{d}_1$ is $(1,1,...,1,0,0,...,0)' = (\iota'_{n_1}, 0')', \mathbf{d}_2 = (0', \iota'_{n_2})'$, and $y = (y'_{n_1}, y'_{n_2})'$

$y = d_1\gamma_1 + d_2\gamma_2 + u = X\alpha + u$, where $X = [d_1 \quad d_2]$. Thus,

$$\hat{\alpha} = \begin{pmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} \iota_{n_1}{}'\iota_{n_1} & 0 \\ 0 & \iota'_{n_2}\iota_{n_2} \end{pmatrix}^{-1} \begin{pmatrix} \iota'_{n_1} y_{n_1} \\ \iota'_{n_2} y_{n_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1}\sum_{n_1} y_i \\ \frac{1}{n_2}\sum_{n_2} y_n \end{pmatrix} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$$

Therefore, $\hat{\gamma}_1, \hat{\gamma}_2$ is the sample mean of the y_i among the men of the sample, and sample mean among the women.

(You can also use 'residual regression'. $\hat{\gamma}_1 = (d_1' M_2 d_1)^{-1}(d_1' M_2 y)$ where $M_2 = I - d_2(d_2'd_2)^{-1}d_2' = I - \frac{1}{n_2}d_2 d_2'$. Since $d_2' d_1 = 0$, $d_1 M_2 = d_1 - \frac{1}{n_2}d_2 d_2' d_1 = d_1$. Thus $\hat{\gamma}_1 = (d_1'd_1)^{-1}(d_1'y) = \frac{1}{n_1}\sum_{i=men} y = \bar{y}_1$

(b) Let $X_1 = [d_1 \quad d_2]$. Then, $y^* = M_{X_1}y, X^* = M_{X_1}X$. In words, this transformation demeans dependent variable(y) among the sample of men and women. In the same way, transformation X to X^* implies, for each column vector(regressor x_k), demean each vector among the sample of men and women, respectively.

(c) $\tilde{\beta} = (X^{*\prime}X^*)^{-1}X^*y^* = (X'M_{X_1}X)^{-1}X'M_{X_1}y$. Also, by residual regression formula, $\hat{\beta} = (X'M_{X_1}X)^{-1}X'M_{X_1}y$. Thus $\hat{\beta} = \tilde{\beta}$ (This is known as Frisch-Waugh-Lovell Theorem)

Exercise 3.15

$$
\begin{aligned}
R^2 = r^2_{y,\hat{y}} \quad &\Leftrightarrow \quad \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{(\sum(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}_i))^2}{\sum(y_i - \bar{y})^2 \sum(\hat{y}_i - \bar{\hat{y}}_i)^2} \\
&\Leftrightarrow \quad \frac{(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y})}{(y - 1\bar{y})'(y - 1\bar{y})} = \frac{\{(\hat{y} - 1\bar{y})'(y - 1\bar{y})\}^2}{(y - 1\bar{y})'(y - 1\bar{y})(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y})} \\
&\Leftrightarrow \quad \{(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y})\}^2 = \{(\hat{y} - 1\bar{y})'(y - 1\bar{y})\}^2 \\
&\Leftrightarrow \quad \{(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y})\}^2 = \{(\hat{y} - 1\bar{y})'(y - \hat{y} + \hat{y} - 1\bar{y})\}^2 \\
&\Leftrightarrow \quad \{(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y})\}^2 = \{(\hat{y} - 1\bar{y})'(\hat{y} - 1\bar{y}) + (\hat{y} - 1\bar{y})'(y - \hat{y})\}^2
\end{aligned}
$$

where $\bar{\hat{y}} = 1/n \sum \hat{y}_i = \bar{y}$ if regressor contains a constant by exercise 3.11. For the last part, $(\hat{y} - 1\bar{y})'(y - \hat{y}) = (P_X y - P_\iota y)'(M_X y) = y' P_X M_X y - y' P_\iota M_X y = 0$ when X has constant regressor. (Note that $P_X = X(X'X)^{-1}X', M_X = I - P_X, P_\iota = 1(1'1)^{-1}1'$, then $P_X M_X = 0$, and $P_\iota M_X = 0$ by using exercise 3.7.) Therefore, equality for last equation always hold when X has constant regressor, and thus $R^2 = r^2_{y,\hat{y}}$

Exercise 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 $SST = \sum_i (y_i - \bar{y}_n)^2$, by adding one regressor the SSR weakly decreases and SST is constant, so R^2 weakly increases. The equality holds when $\widehat{\beta}_2 = 0$.

A more detailed solution would be: let $X = [X_1 X_2]$, then $R_1^2 = 1 - \frac{\tilde{e}'\tilde{e}}{(y-1\bar{y})'(y-1\bar{y})} = 1 - \frac{y' M_{X_1} y}{(y-1\bar{y})'(y-1\bar{y})}$. By the similar calculation, $R_2^2 = 1 - \frac{y' M_X y}{(y-1\bar{y})'(y-1\bar{y})}$ (Again, note that P_X, P_{X_1} are projection matrix with X, X_1 respectively, and M_X, M_{X_1} are orthogonal projection matrix)

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

Note that $(P_X - P_{X_1})^2 = P_X^2 - P_X P_{X_1} - P_{X_1} P_X + P_{X_1}^2 = P_X - P_{X_1}$, by using exercise 3.7.($P_X P_{X_1} = P_X X_1 (X_1' X_1)^{-1} X_1' = X_1 (X_1' X_1)^{-1} X_1' = P_{X_1}$). Thus $(P_X - P_{X_1})$ is idempotent. You can also easily check $(P_X - P_{X_1})$ is symmetric. Since symmetric idempotent matrix's characteristic roots are either 0 or 1, thus it is positive semi-definite. Thus $y'(P_X - P_{X_1})y = ||(P_X - P_{X_1})y||^2 \geq 0$ holds for any y.

Equality holds when $(P_X - P_{X_1})y = 0$ for some y. This holds when X_2 is orthogonal to X_1 and y, i.e., $X_1' X_2 = 0 \& X_2' y = 0$. To see this, note that $P = P_1 + P_2$ when $X_1' X_2 = 0$ by exercise 3.10. Thus

$$
(P_X - P_{X_1})y = P_{X_2} y = X_2 (X_2' X_2)^{-1} X_2' y = 0
$$

In this case, adding X_2 variables does not help to increase R^2

(If we allow multi-collinearity of the regressors, another obvious case will be when column vector of X_2 are all linear combination of X_1, i.e. $P_X = P_{X_1}$)

Exercise 3.17

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(1-h_{ii})^{-2}\hat{e}_i^2 \geq \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 = \hat{\sigma}^2$$

where, inequality comes from the fact $0 \leq h_{ii} \leq 1$. Thus equality $\tilde{\sigma}^2 = \hat{\sigma}^2$ holds when $h_{ii} = 0 \quad \forall i$. This is not possible because $\sum_i h_{ii} = k$. However, difference between $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ goes to 0 as $h_{ii} \to 0$ when $n \to \infty$

Exercise 3.19

1. With restricted sample, we get the same estimates for log wage regression as in equation (3.43) (Let's defer standard error calculations to next chapter's exercises)

Table 1: OLS Estimates Result for Wage Regression:
All Single Asian Males with Less than 45 Years of Experience ($n = 267$)

Variables	ln(Wage per hours)
Education	0.144
Experience	0.043
Experience squared/100	-0.095
Constant	0.531
R^2	0.389
SSE(Sum of Squared Errors)	82.505
observations	267

Let y_i be log wages and x_i be years of education(x_{1i}), experience (x_{2i}), experience^2/100 (x_{3i}), and an intercept. Note that $R^2 = 1 - \frac{y'My}{y'M_1y}$ and $SSE = y'My$ where $M = I - X(X'X)^{-1}X', M_1 = I - 1(1'1)^{-1})1'$

2.

The coefficient of residual regression and the coefficient of education on wages in the previous regression in part 1 are the same. We know that they should be same by Frisch-Waugh-Lovell theorem.

3.

SSE are the same for both regression estimates. However, R^2 is different. To see this, let $X_1 = [x_1], X_2 = [1 \quad x_2 \quad x_3]$, which is $n \times 1, n \times 3$ matrix, respectively. In the previous problem we report estimates from $y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}$. Let $\tilde{e}, \bar{e}$ are residuals from regress y on X_2, and from regress X_1 on X_2, respectively. Note that we report OLS estimates from $\tilde{e}$ on $\bar{e}$ in the table 2.

Table 2: OLS Coefficient Estimates for Education on Wage:Using Residual Regression Approach

Variables	First Step Residual
Coefficient estimates	
Second Step Residual	0.144
R^2	0.369
SSE	82.505
observations	267

Dependent variable "First Step Residual" is the residual from regression $log(Wage)$ on experience,experience^2/100, and a constant. Similarly, "Second Step Residual" can be obtained from regression of education on experience, experience^2/100, and a constant.

Residual from residual regression $\hat{\hat{e}}$ should be same with residual from original equation $\hat{e}$ because;

$$\begin{aligned} \hat{\hat{e}} = \tilde{e} - \bar{e}\tilde{\beta}_1 &= M_{X_2}y - M_{X_2}X_1\tilde{\beta}_1 \\ &= M_{X_2}y - M_{X_2}X_1\hat{\beta}_1 (\text{By F-W-L theorem} \hat{\beta}_1 = \tilde{\beta}_1) \\ &= M_{X_2}(X_2\hat{\beta}_2 + \hat{e}) = M_{X_2}\hat{e} = (I - P_{X_2})\hat{e} = \hat{e} \end{aligned}$$

where the last equality holds since $X_2'\hat{e} = 0$ by exercise 3.4. Thus sum of square errors should be same.

R^2 from original OLS and residual regression are as follows;

$$\begin{aligned} R^2_{original} &= 1 - \frac{\hat{e}'\hat{e}}{(y - 1\bar{y})'(y - 1\bar{y})} = 1 - \frac{\hat{e}'\hat{e}}{y'M_1y} \\ R^2_{residual} &= 1 - \frac{\hat{\hat{e}}'\hat{\hat{e}}}{(\tilde{e} - 1\bar{\tilde{e}})'(\tilde{e} - 1\bar{\tilde{e}})} = 1 - \frac{\hat{e}'\hat{e}}{y'M_{X_2}M_1M_{X_2}y} = 1 - \frac{\hat{e}'\hat{e}}{y'M_{X_2}y} \end{aligned}$$

since $\hat{\hat{e}} = \hat{e}, \tilde{e} = M_{X_2}y$, and $M_{X_2}M_1 = M_{X_2}$. By the similar reason in exercise 3.16, $y'M_1y - y'M_{X_2}y = y'(M_1 - M_{X_2})y \geq 0$ since $M_1 - M_{X_2}$ are idempotent. Therefore, $y'M_1y \geq y'M_{X_2}y$, and $R^2_{original} \geq R^2_{residual}$

Exercise 3.20

(a) $\sum_{i=1}^n \hat{e}_i = 7.994 \times 10^{-14}$
(b) $\sum_{i=1}^n x_{1i}\hat{e}_i = \mathbf{x}_1'\hat{e} = -4.181 \times 10^{-12}$
(c) $\sum_{i=1}^n x_{2i}\hat{e}_i = \mathbf{x}_2'\hat{e} = -2.751 \times 10^{-12}$
(d) $\sum_{i=1}^n x_{1i}^2\hat{e}_i = \mathbf{x}_1^{2\prime}\hat{e} = 133.133$
(e) $\sum_{i=1}^n x_{2i}^2\hat{e}_i = \mathbf{x}_2^{2\prime}\hat{e} = 1.035 \times 10^{-11}$
(f) $\sum_{i=1}^n \hat{y}_i\hat{e}_i = \hat{y}'\hat{e} = -7.002 \times 10^{-13}$
(g) $\sum_{i=1}^n \hat{e}_i^2 = \hat{e}'\hat{e} = 82.505$

(a), (b), (c), (e), and (f) are close to 0, which is consistent with the theoretical properties of OLS, as we did in exercise 3.3. Note that numerical calculations do not return exact zeros.