Problem Set #4 (2^{nd} Half) Solutions

Economics 709 **Fall 2020**

7.28 (a) Here are the OLS estimates and the corresponding robust standard errors:

Variables	Coeff Est	Robust Std Err
Education	0.144	(0.012)
Experience	0.043	(0.012)
Experience squared/100	-0.095	(0.034)
Constant	0.531	(0.20)
R^2	0.389	
observations	267	

(b) Set $exp = 10$. Then,

$$\theta = g(\beta) = \frac{\beta_1}{\beta_2 + \frac{\beta_3}{50}(10)} = \frac{\beta_1}{\beta_2 + \frac{\beta_3}{5}}$$

So,

$$\hat{\theta} = g(\hat{\beta}) = \frac{\hat{\beta}_1}{\hat{\beta}_2 + \frac{\hat{\beta}_3}{5}} = \frac{0.144}{0.043 - 0.019} = 6.1$$

(c) The asymptotic standard error for $\hat{\theta}$ is $s(\hat{\theta}) = \sqrt{\hat{G}'\hat{V}_\beta\hat{G}}$, where $\hat{V}_\beta$ is the estimated asymptotic robust variance-covariance matrix that is used to generate the standard errors in (a). And, $\hat{G}$ is a consistent estimate of

$$G = \frac{\partial}{\partial \beta} g(\beta) = \begin{pmatrix} \frac{1}{\beta_2 + \frac{\beta_3}{50} exp} \\ -\frac{\beta_1}{(\beta_2 + \frac{\beta_3}{50} exp)^2} \\ -\frac{\frac{\beta_1}{50} exp}{(\beta_2 + \frac{\beta_3}{50} exp)^2} \end{pmatrix}$$

To obtain $\hat{G}$ simply plug in the OLS estimates for β and set experience level to 10. Here are the estimates and standard errors for a few experience levels:

$\hat{\theta}$	$s(\hat{\theta})$
6.11	(1.63)

(d) A 90% CI for θ is: $\hat{\theta} \pm 1.645 s(\hat{\theta})$:

90% CI
[3.43 , 8.79]

8.1 $\tilde{\beta}$ solves

$$\min_{\beta:\beta_2=0} (Y - X_1\beta_1 - X_2\beta_2)'(Y - X_1\beta_1 - X_2\beta_2)$$

The Lagrangian is:

$$\mathcal{L} = (Y - X_1\beta_1 - X_2\beta_2)'(Y - X_1\beta_1 - X_2\beta_2) + \lambda'(\beta_2 - 0)$$

FOCs:

$$\begin{aligned}
\left[\frac{\partial \mathcal{L}}{\partial \beta_1}\right] \qquad 0 &= -2X_1'(Y - X_1\tilde{\beta}_1 - X_2\tilde{\beta}_2) \\
\left[\frac{\partial \mathcal{L}}{\partial \lambda}\right] \qquad 0 &= \tilde{\beta}_2
\end{aligned}$$

Substituting,

$$0 = -2X_1'(Y - X_1\tilde{\beta}_1) \quad \Longrightarrow \quad \tilde{\beta}_1 = \left(X_1'X_1\right)^{-1} X_1'Y$$

8.3 $\tilde{\beta}$ solves

$$\min_{\beta:\beta_1+\beta_2=0} (Y - X_1\beta_1 - X_2\beta_2)' (Y - X_1\beta_1 - X_2\beta_2)$$

The Lagrangian is:

$$\mathcal{L} = (Y - X_1\beta_1 - X_2\beta_2)' (Y - X_1\beta_1 - X_2\beta_2) + \lambda'(\beta_1 + \beta_2 - 0)$$

FOCs:

$$\begin{aligned}
\left[\frac{\partial \mathcal{L}}{\partial \beta_1}\right] \qquad 0 &= -2X_1'(Y - X_1\tilde{\beta}_1 - X_2\tilde{\beta}_2) + \lambda \\
\left[\frac{\partial \mathcal{L}}{\partial \beta_2}\right] \qquad 0 &= -2X_2'(y - X_1\tilde{\beta}_1 - X_2\tilde{\beta}_2) + \lambda \\
\left[\frac{\partial \mathcal{L}}{\partial \lambda}\right] \qquad 0 &= \tilde{\beta}_1 + \tilde{\beta}_2
\end{aligned}$$

Substituting,

$$0 = -2(X_1 - X_2)' \left[Y - (X_1 - X_2)\tilde{\beta}_1\right] \implies \tilde{\beta}_1 = -\tilde{\beta}_2 = \left((X_1 - X_2)' (X_1 - X_2)\right)^{-1} (X_1 - X_2)' Y$$

8.4 (a) The restricted model is $Y_i = \alpha + e_i$. The CLS estimator is $\tilde{\alpha} = \frac{1}{n}\sum_{i=1}^n Y_i$.

8.22 (a) We can rewrite the restriction as $\beta_1 = 2\beta_2$. Substituting this into the equation we find

$$Y_i = (2x_{1i} + x_{2i})\beta_2 + e_i$$

The CLS estimate of β_2 is the simple regression

$$\tilde{\beta}_2 = \frac{\sum_{i=1}^n (2x_{1i} + x_{2i})Y_i}{\sum_{i=1}^n (2x_{1i} + x_{2i})^2}$$

and that for β_1 is

$$\tilde{\beta}_1 = 2\tilde{\beta}_2 = 2\frac{\sum_{i=1}^n (2x_{1i} + x_{2i})Y_i}{\sum_{i=1}^n (2x_{1i} + x_{2i})^2}.$$

These expressions can also be derived directly through the Lagrangian with constraint $\frac{\beta_1}{\beta_2} = 2$.

(b) By the WLLN and CLT,

$$\sqrt{n}(\tilde{\beta}_1 - \beta_1) = 2\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (2x_{1i} + x_{2i})e_i}{\frac{1}{n}\sum_{i=1}^n (2x_{1i} + x_{2i})^2} \xrightarrow{d} N\left(0, \frac{E((2x_{1i} + x_{2i})^2 e_i^2)}{(E(2x_{1i} + x_{2i})^2)^2}\right).$$

9.1 Partition $X = [W\ \ Z]$, where $Z = X_{k+1}$, and similarly partition $\beta = [\gamma'\ \ \theta]'$, where $\theta = \beta_{k+1}$ is scalar. Unrestricted OLS y on X yields $\hat{\beta} = (X'X)^{-1}X'y$ and $\hat{\varepsilon} = y - X\hat{\beta}$. The restriction to impose is $\theta = \beta_{k+1} = 0$. Recall that restricted LS $(\tilde{\beta})$ is related to unrestricted OLS $(\hat{\beta})$:

$$\begin{aligned}
\tilde{\beta} &= \hat{\beta} - (X'X)^{-1}[0_k\ \ 1]'\left([0_k\ \ 1](X'X)^{-1}[0_k\ \ 1]'\right)^{-1}\underbrace{[0_k\ \ 1]\hat{\beta}}_{\hat{\theta}} \\
&= \hat{\beta} - (X'X)^{-1}[0_k\ \ 1]'\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\theta} \qquad (1)
\end{aligned}$$

The restricted LS residual is

$$\tilde{\varepsilon} = y - X\tilde{\beta} = y - X\hat{\beta} - X(\tilde{\beta} - \hat{\beta}) = \hat{\varepsilon} - X(\tilde{\beta} - \hat{\beta})$$

Since $X'\hat{\varepsilon} = 0$ and using (1),

$$\begin{aligned}
\tilde{\varepsilon}'\tilde{\varepsilon} &= \hat{\varepsilon}'\hat{\varepsilon} + (\tilde{\beta} - \hat{\beta})'(X'X)(\tilde{\beta} - \hat{\beta}) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\theta}\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}[0_k\ \ 1](X'X)^{-1}(X'X)(X'X)^{-1}[0_k\ \ 1]'\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\theta} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\theta}\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}[0_k\ \ 1](X'X)^{-1}[0_k\ \ 1]'\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\theta} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\theta}\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}[(X'X)^{-1}]_{k+1,k+1}\left([(X'X)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\theta} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \frac{\hat{\theta}^2}{[(X'X)^{-1}]_{k+1,k+1}}
\end{aligned}$$

Now let $\bar{R}^2_{k+1}$ and $\bar{R}^2_k$ be the adjusted R-squared for unrestricted and restricted LS regressions. Let $SST = \frac{1}{n-1}\sum_i (y_i - \bar{y})^2$, and note $s^2 = \frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon}$.

$$\begin{aligned}
\bar{R}^2_{k+1} > \bar{R}^2_k &\iff 1 - \frac{\frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon}}{SST} > 1 - \frac{\frac{1}{n-k}\tilde{\varepsilon}'\tilde{\varepsilon}}{SST} \iff \frac{1}{n-k}\tilde{\varepsilon}'\tilde{\varepsilon} > \frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon} \\
&\iff (n-k-1)(\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}) > \hat{\varepsilon}'\hat{\varepsilon} \iff \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon}} > 1 \\
&\iff \frac{\hat{\theta}^2}{s^2[(X'X)^{-1}]_{k+1,k+1}} > 1 \iff \frac{\hat{\beta}^2_{k+1}}{s(\hat{\beta}_{k+1})^2} > 1 \iff \left|\frac{\hat{\beta}_{k+1}}{s(\hat{\beta}_{k+1})}\right| > 1 \\
&\iff |T_{k+1}| > 1
\end{aligned}$$

9.2 (a) By the asymptotic properties of the OLS estimator for each sample, we know that

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) &\xrightarrow{d} N\left(0, V_1\right) \\
\sqrt{n}\left(\hat{\beta}_2 - \beta_2\right) &\xrightarrow{d} N\left(0, V_2\right),
\end{aligned}$$

where

$$\begin{aligned}
V_1 &= \mathbb{E}\left[x_{1i}x_{1i}'\right]^{-1}\mathbb{E}\left[x_{1i}x_{1i}'e_{1i}^2\right]\mathbb{E}\left[x_{1i}x_{1i}'\right]^{-1} \\
V_2 &= \mathbb{E}\left[x_{2i}x_{2i}'\right]^{-1}\mathbb{E}\left[x_{2i}x_{2i}'e_{2i}^2\right]\mathbb{E}\left[x_{2i}x_{2i}'\right]^{-1}.
\end{aligned}$$

Using matrix notation, we may write

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_1-\beta_1 \\ \hat{\beta}_2-\beta_2 \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{n}\sum_{i=1}^n x_{i1}x_{i1}'\right)^{-1} & 0 \\ 0 & \left(\frac{1}{n}\sum_{i=1}^n x_{i1}x_{i1}'\right)^{-1} \end{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} x_{i1}e_{i1} \\ x_{i2}e_{i2} \end{pmatrix}.$$

Independence between sample 1 and 2 and CLT implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{pmatrix} x_{i1}e_{i1} \\ x_{i2}e_{i2} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbb{E}\left[x_{1i}x_{1i}'e_{1i}^2\right] & 0 \\ 0 & \mathbb{E}\left[x_{2i}x_{2i}'e_{2i}^2\right] \end{pmatrix}\right).$$

Thus, LLN and Slutsky's lemma implies that,

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_1-\beta_1 \\ \hat{\beta}_2-\beta_2 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}\right).$$

By Continuous Mapping Theorem (CMT),

$$\sqrt{n}\left(\left(\hat{\beta}_1-\hat{\beta}_2\right)-\left(\beta_1-\beta_2\right)\right) \xrightarrow{d} N\left(0, V_1+V_2\right).$$

(b) From (a), an appropriate test statistic for $H_0 : \beta_2 = \beta_1$ is the Wald statistic.

$$W_n = n\left(\hat{\beta}_1-\hat{\beta}_2\right)'\left(\hat{V}_1+\hat{V}_2\right)^{-1}\left(\hat{\beta}_1-\hat{\beta}_2\right),$$

where $\hat{V}_1$ and $\hat{V}_2$ are consistent estimators of V_1, V_2.

(c) From (a) and (b),

$$\sqrt{n}\left(\left(\hat{\beta}_1-\hat{\beta}_2\right)-\left(\beta_1-\beta_2\right)\right) \xrightarrow{d} N\left(0, V_1+V_2\right),$$

and $\hat{V}_1 \xrightarrow{p} V_1$, $\hat{V}_2 \xrightarrow{p} V_2$. Under H_0, $W_n \xrightarrow{d} \chi_k^2$ as $n \to \infty$.

9.4 (a) $Pr(Reject|H_0) = Pr(W < c_1|H_0) + Pr(W > c_2|H_0) \xrightarrow{p} \frac{\alpha}{2} + 1 - (1-\frac{\alpha}{2}) = \alpha$ since $W \xrightarrow{d} \chi_q^2$ under H_0. Thus the asymptotic size of the test is α.

(b) This is not a good test of H_0 versus H_1. For W close to zero, it means that $\hat{\theta}$ is close to zero, and it is more likely that the true value θ is close to zero. Thus, H_0 is more likely to be true and we are supposed to accept H_0, instead of rejecting it. This proposed decision rule would lead to a loss in power.

9.7 The expected wage for a 40-year-old worker is $E(y_i|x_i = 40) = 40\beta_1 + 1600\beta_2$ because $E(e_i|x_i) = 0$. Let $\theta = 40\beta_1 + 1600\beta_2 - 20$. The test hypotheses are H_0: $\theta = 0$, H_1: $\theta \neq 0$. Since θ is a scalar, we can use t-test here. The asymptotic distribution of $\hat{\theta} = 40\hat{\beta}_1 + 1600\hat{\beta}_2 - 20$ is

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \begin{pmatrix} 40 & 1600 \end{pmatrix} V_\beta \begin{pmatrix} 40 \\ 1600 \end{pmatrix}\right) \sim N(0, V_\theta)$$

where V_β is the asymptotic variance of $(\hat{\beta}_1, \hat{\beta}_2)'$, $V_\theta = 1600V_\beta^{11} + 64000V_\beta^{21} + 64000V_\beta^{12} + 2560000V_\beta^{22}$ is the asymptotic variance of θ. Thus, the standard error of $\hat{\theta}$ is $se(\hat{\theta}) = \sqrt{\frac{1}{n}\hat{V}_\theta}$, where $\hat{V}_\theta$ is constructed from the White estimator $\hat{V}_\beta$ according to the above formula.

The t-statistic is $t = \frac{\hat{\theta}-0}{se(\hat{\theta})}$. Under H_0, $t \xrightarrow{d} N(0,1)$. The decision rule of our test is: reject H_0 if $|t| > q_{1-\frac{\alpha}{2}}$, accept H_0 if $|t| \leq q_{1-\frac{\alpha}{2}}$, where $q_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-th quantile of $N(0,1)$ and α is the pre-specified asymptotic size of the test.