Problem Set #4

Danny Edgel
Econ 709: Economic Statistics and Econometrics I
Fall 2020

December 8, 2020

Collaborated with Sarah Bass, Emily Case, Michael Nattinger, and Alex Von Hafften

Question 1

(a) The table below displays the coefficient estimates, alongside robust standard errors.

VARIABLES	(1) log(wage)
Education	0.144***
	(0.0118)
Experience	0.0426***
	(0.0125)
Experience^2	-0.0951***
	(0.0341)
Constant	0.531***
	(0.202)
Observations	267
R-squared	0.389
Sum-of-squared Errors	82.50

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(b) In terms of the model parameters, with $experience = 10$,

$$\theta = \frac{\beta_1}{\beta_2 + \frac{1}{5}\beta_3}$$

Using the parameter estimates from the model,

$$\hat{\theta} \approx 6.109$$

(c) The asymptotic standard error of $\hat{\theta}$ is the square root of its asymptotic variance. Since θ is a function of β, we can use the delta method to solve for the variance of $\hat{\theta}$ as a function of the variance-covariance matrix of $\hat{\beta}$:

$$\sqrt{n}\left(\hat{\theta}-\theta\right) \rightarrow_d f'(\hat{\beta})\mathcal{N}(0,V) \equiv \mathcal{N}\left(0, f'(\hat{\beta})'Vf'(\hat{\beta})\right)$$

Where V is the variance-covariance matric of $\hat{\beta}$ and

$$f'(\hat{\beta}) = \begin{pmatrix} \frac{\partial f(\hat{\beta})}{\partial \hat{\beta}_1} \\ \frac{\partial f(\hat{\beta})}{\partial \hat{\beta}_2} \\ \frac{\partial f(\hat{\beta})}{\partial \hat{\beta}_3} \\ \frac{\partial f(\hat{\beta})}{\partial \hat{\beta}_4} \end{pmatrix} = \begin{pmatrix} \frac{1}{\hat{\beta}_2+\frac{1}{5}\hat{\beta}_3} \\ -\frac{\hat{\beta}_1}{\left(\hat{\beta}_2+\frac{1}{5}\hat{\beta}_3\right)^2} \\ -\frac{\hat{\beta}_1}{5\left(\hat{\beta}_2+\frac{1}{5}\hat{\beta}_3\right)^2} \\ 0 \end{pmatrix}$$

(d) Using the results from the regression summarized in part (a),

$$s(\hat{\theta}) \approx 1.63$$
$$90\% \text{ c.i.} = [\hat{\theta}-1.645s(\hat{\theta}), \hat{\theta}+1.645s(\hat{\theta})] \approx [3.428, 8.790]$$

Question 2

According to equation (8.3),

$$\widetilde{\beta}_{CLS} = \arg\min_{R'\beta=c} \text{SSE}(\beta)$$

Where $R = \begin{pmatrix} 0 \\ I_{k_2} \end{pmatrix}$ and $c = 0$. Then, (8.3) can be simplified as the following unconstrained optimization problem:

$$\widetilde{\beta}_{CLS} = \text{argmin SSE}\begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}$$

WAnd where $\hat{\beta}_{OLS}$ from the regression of Y on X_1 is defined as:

$$\hat{\beta}_{OLS} = \text{argmin SSE}(\beta_1)$$

Question 3

By equation (8.3),

$$\widetilde{\beta}_{CLS} = \arg\min_{R'\beta=c} \text{SSE}(\beta)$$

Where, $\text{SSE}(\beta) = (Y - X\beta)'(Y - X\beta) = (Y - X_1\beta_1 - X_2\beta_2)'(Y - X_1\beta_1 - X_2\beta_2)$ and, in this case, $R = \begin{pmatrix} I_k \\ I_k \end{pmatrix}$ and $c = 0$. Then,

$$\begin{aligned}
\widetilde{\beta}_{CLS} &= \arg\min_{R'\beta=c} (Y - X\beta)'(Y - X\beta) \\
\mathscr{L} &= (Y - X_1\beta_1 - X_2\beta_2)'(Y - X_1\beta_1 - X_2\beta_2) - \lambda(\beta_1 + \beta_2) \\
\frac{\partial \mathscr{L}}{\partial \beta_1} &= -2X_1'(Y - X_1\beta_1 - X_2\beta_2) - \lambda = 0 \\
\frac{\partial \mathscr{L}}{\partial \beta_2} &= -2X_2'(Y - X_1\beta_1 - X_2\beta_2) - \lambda = 0 \\
\frac{\partial \mathscr{L}}{\partial \lambda} &= \beta_1 + \beta_2 = 0 \\
\beta_1 &= -\beta_2 \\
-2X_1'(Y - X_1\beta_1 - X_2\beta_2) &= -2X_2'(Y - X_1\beta_1 - X_2\beta_2) \\
-2X_1'(Y + X_1\beta_2 - X_2\beta_2) &= -2X_2'(Y + X_1\beta_2 - X_2\beta_2) \\
-2X_1'Y - 2X_1'(X_1 - X_2)\beta_2 &= -2X_2'Y - 2X_2'(X_1 - X_2)\beta_2 \\
2(X_2 - X_1)'(X_1 - X_2)\beta_2 &= 2(X_1 - X_2)'Y \\
\beta_2 &= [(X_2 - X_1)'(X_1 - X_2)]\,(X_1 - X_2)'Y \\
\beta_1 &= [(X_2 - X_1)'(X_2 - X_1)]\,(X_2 - X_1)'Y
\end{aligned}$$

Thus,

$$\widetilde{\beta}_{CLS} = \begin{pmatrix} [(X_2 - X_1)'(X_2 - X_1)]\,(X_2 - X_1)'Y \\ [(X_2 - X_1)'(X_1 - X_2)]\,(X_1 - X_2)'Y \end{pmatrix}$$

Question 4

The linear projection model $Y = \alpha + X\beta + \varepsilon$ can be written as $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, where X_1 is a vector of ones. We showed in question 1 that the CLS estimate of this specification with $\beta_2 = 0$ is simply the OLS estimate of Y on X_1. We've shown in prior problem sets that the OLS estimate of Y on a constant is $\mathbb{E}Y$, So the $\widetilde{\alpha}_{CLS} = \mathbb{E}Y$.

Question 5

8.22 The proposed restriction on β can be written as $r(\beta) = 0$, where $r(\beta) = \frac{\beta_1}{\beta_2} - 2$.

(a) The CLS estimator for this restriction is defined as:

$$\widetilde{\beta}_{CLS} = \arg\min_{r(\beta)=0} \text{SSE}(\beta)$$

There is no closed-form solution for this estimator, but we can rewrite this specific restriction as $\beta_1 = 2\beta_2$, which gives us the specification $Y = (2X_1 + X_2)\beta_2 + \varepsilon$.

The estimator for this specification is:

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

Which can be plugged back into the constraint to retrieve $\widetilde{\beta}_1$, ultimately yielding the estimator:

$$\widetilde{\beta}_{CLS} = \begin{pmatrix} 2\frac{\sum_{i=1}^{n}(2x_{1i}+x_{2i})y_i}{\sum_{i=1}^{n}(2x_{1i}+x_{2i})^2} \\ \frac{\sum_{i=1}^{n}(2x_{1i}+x_{2i})y_i}{\sum_{i=1}^{n}(2x_{1i}+x_{2i})^2} \end{pmatrix}$$

(b) If this restriction is true, then $\widetilde{\beta}_1 \to_p \beta_1$, and, by the CLT,

$$\sqrt{n}(\widetilde{\beta}_1 - \beta_1) = 2\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(2x_{1i}+x_{2i})\varepsilon_i}{\frac{1}{n}\sum_{i=1}^{n}(2x_{1i}+x_{2i})^2} \to_d \mathcal{N}(0, V)$$

Where

$$V = \frac{\mathbb{E}\,(2x_{1i}+x_{2i})^2\varepsilon_i^2}{\left(\mathbb{E}\,(2x_{1i}+x_{2i})^2\right)^2} = \frac{\mathbb{E}\,(2x_{1i}+x_{2i})^2}{\left(\mathbb{E}\,(2x_{1i}+x_{2i})^2\right)^2}\sigma^2$$

Question 6

9.1 Let $\beta = [\phi\ \beta_{k+1}]$ represent the OLS coefficients from a partitioned regression of Y on $Z = [X\ X_{k+1}]$. Then,

$$\hat{\beta} = (Z'Z)^{-1}Z'Y$$

Now, let $RZ = c$ represent a restriction on β where $R = \begin{pmatrix} 0_k \\ 1 \end{pmatrix}$ and $c = 0$.

Then, by equation (8.8),

$$\begin{aligned}\widetilde{\beta} &= \hat{\beta} - (Z'Z)^{-1}R\left[R'(Z'Z)^{-1}R\right]^{-1}R'\hat{\beta} \\ &= \hat{\beta} - (Z'Z)^{-1}\begin{pmatrix} 0_k \\ 1 \end{pmatrix}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1}\end{aligned}$$

Where $\widetilde{\beta}$ has the residual

$$\widetilde{\varepsilon} = Y - Z\widetilde{\beta} = Y - Z\hat{\beta} + Z(Z'Z)^{-1}\begin{pmatrix} 0_k \\ 1 \end{pmatrix}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1} = Y - Z\hat{\beta} - Z(\widetilde{\beta} - \hat{\beta})$$
$$= \hat{\varepsilon} - Z(\widetilde{\beta} - \hat{\beta})$$

Then, we can calculate:

$$\begin{aligned}
\widetilde{\varepsilon}'\widetilde{\varepsilon} &= (\hat{\varepsilon} - Z(\widetilde{\beta} - \hat{\beta}))'(\hat{\varepsilon} - Z(\widetilde{\beta} - \hat{\beta})) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + (\widetilde{\beta} - \hat{\beta})'(Z'Z)(\widetilde{\beta} - \hat{\beta}) \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\beta}_{k+1}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\begin{pmatrix}0_k \\ 1\end{pmatrix}(Z'Z)^{-1}(Z'Z)(Z'Z)^{-1}\begin{pmatrix}0_k \\ 1\end{pmatrix}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\beta}_{k+1}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\begin{pmatrix}0_k \\ 1\end{pmatrix}(Z'Z)^{-1}\begin{pmatrix}0_k \\ 1\end{pmatrix}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\beta}_{k+1}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}[(Z'Z)^{-1}]_{k+1,k+1}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1} \\
&= \hat{\varepsilon}'\hat{\varepsilon} + \hat{\beta}_{k+1}\left([(Z'Z)^{-1}]_{k+1,k+1}\right)^{-1}\hat{\beta}_{k+1}
\end{aligned}$$

Since $\hat{\beta}_{k+1}$ and $[(Z'Z)^{-1}]_{k+1,k+1}$ are scalars,

$$\widetilde{\varepsilon}'\widetilde{\varepsilon} = \hat{\varepsilon}'\hat{\varepsilon} + \frac{\hat{\beta}^2_{k+1}}{[(Z'Z)^{-1}]_{k+1,k+1}}$$

Then, letting $s^2 = \frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon}$,we can identify the condition for the adjusted R^2 of the unrestricted model being higher than that of the restricted model and solve:

$$\begin{aligned}
1 - \frac{\frac{1}{n-k-1}\hat{\varepsilon}'\hat{\varepsilon}}{\frac{1}{n-1}\sum_{i=1}^n(y_i - \overline{y})^2} &> 1 - \frac{\frac{1}{n-k}\widetilde{\varepsilon}'\widetilde{\varepsilon}}{\frac{1}{n-1}\sum_{i=1}^n(y_i - \overline{y})^2} \\
-(n-k)\hat{\varepsilon}'\hat{\varepsilon} &> -(n-k-1)\widetilde{\varepsilon}'\widetilde{\varepsilon} \\
(n-k)\hat{\varepsilon}'\hat{\varepsilon} &< (n-k-1)\left(\hat{\varepsilon}'\hat{\varepsilon} + \frac{\hat{\beta}^2_{k+1}}{[(Z'Z)^{-1}]_{k+1,k+1}}\right) \\
\hat{\varepsilon}'\hat{\varepsilon} &< (n-k-1)\frac{\hat{\beta}^2_{k+1}}{[(Z'Z)^{-1}]_{k+1,k+1}} \\
\frac{\hat{\beta}^2_{k+1}}{s^2[(Z'Z)^{-1}]_{k+1,k+1}} = \left| \frac{\hat{\beta}_{k+1}}{s\left(\hat{\beta}_{k+1}\right)} \right| &> 1 \\
\left| T_{k+1} \right| > 1 \ \blacksquare
\end{aligned}$$

9.2 (a) Since $\mathbb{E}\,X_1e_1 = \mathbb{E}\,X_2e_2 = 0$, we know

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \to_d \mathcal{N}\left(0, \mathbb{E}\,x'_{ji}x_{ji}^{-1}\,\mathbb{E}\,x'_{ji}x_{ji}e^2_{ji}\,\mathbb{E}\,x'_{ji}x_{ji}^{-1}\right)$$

For $j = 1, 2$. Since the two samples are independent, $Cov(x_{1i}, x_{2i}) = Cov(e_{1i}, e_{2i}) = 0$, so

$$\begin{aligned}
\sqrt{n}\begin{pmatrix}\hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2\end{pmatrix} &= \begin{pmatrix}\mathbb{E}\,x'_{1i}x_{1i}^{-1} & 0 \\ 0 & \mathbb{E}\,x'_{2i}x_{2i}^{-1}\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n\begin{pmatrix}x_{1i} \\ x_{2i}\end{pmatrix} \\
&\to_d \mathcal{N}\left(0, \begin{pmatrix}\mathbb{E}\,x'_{1i}x_{1i}^{-1}\,\mathbb{E}\,x'_{1i}x_{1i}e^2_{1i}\,\mathbb{E}\,x'_{1i}x_{1i}^{-1} & 0 \\ 0 & \mathbb{E}\,x'_{2i}x_{2i}^{-1}\,\mathbb{E}\,x'_{2i}x_{2i}e^2_{2i}\,\mathbb{E}\,x'_{2i}x_{2i}^{-1}\end{pmatrix}\right)
\end{aligned}$$

Then, we can solve,

$$\sqrt{n}\left((\hat\beta_2-\hat\beta_1)-(\beta_2-\beta_1)\right)=\sqrt{n}\left((\hat\beta_2-\beta_2)-(\hat\beta_1-\beta_1)\right)$$
$$\to_d \mathcal{N}\left(0, \mathbb{E}\, x'_{1i}x_{1i}{}^{-1}\, \mathbb{E}\, x'_{1i}x_{1i}e^2_{1i}\, \mathbb{E}\, x'_{1i}x_{1i}{}^{-1} + \mathbb{E}\, x'_{2i}x_{2i}{}^{-1}\, \mathbb{E}\, x'_{2i}x_{2i}e^2_{2i}\, \mathbb{E}\, x'_2\right.$$

(b) Let $\theta=\beta_2-\beta_1$. By equation (9.6), an appropriate Wald statistic is

$$W=\hat\theta'\hat V\theta^{-1}\hat\theta$$

Then, given our result from (a),

$$W=(\hat\beta_2-\hat\beta_1)'\left(\hat V_1+\hat V_2\right)(\hat\beta_2-\hat\beta_1)$$

Where $\hat V_i$ is a consistent estimator for $\mathbb{E}\, x'_{ji}x_{ji}{}^{-1}\, \mathbb{E}\, x'_{ji}x_{ji}e^2_{ji}\, \mathbb{E}\, x'_{ji}x_{ji}{}^{-1}$.

(c) Since $\hat\beta_2-\hat\beta_1\to_p\beta_2-\beta_1$ and $\hat V_1+\hat V_2\to_p V_1+V_2$, under our null hypothesis,

$$W\to_d\chi^2_k$$

Question 7

9.4

(a) The size of a test is equal to the probability of rejection. Then,

$$Pr(Reject|H_0)=Pr(W<c_1|H_0)+Pr(W>C_2|H_0)\to_p\frac{\alpha}{2}+(1-(1-\frac{\alpha}{2}))=\frac{\alpha}{2}+\frac{\alpha}{2}=\alpha$$

(b) This is not at all a good test of the null hypothesis, beacuse the lower-tail rejection standard, $W<c_1$, includes both the true value of the estimator under the null hypothesis and a fat section of the χ^2 distribution. As a result, this test is extremely weak and will result in many rejections of true null hypotheses.

Question 8

Our null hypothesis is $H_0: 40\beta_1+40^2\beta_2=20$, so to test this hypothesis, we would acquire estimator $\hat\beta=\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}$, then construct a consistent estimator for its asymptotic variance, $\hat{V}_{\hat{\beta}}$ to acquire a test statistic: $t = \frac{\hat{\theta}}{se(\hat{\theta})}$, where:

$$\hat{\theta} = 40\hat{\beta}_1 + 40^2\hat{\beta}_2 - 20$$

$$\text{Under } H_0: \ \sqrt{n}(\hat{\theta} - 0) \to_d \mathcal{N}\left(0, \begin{pmatrix} 40 \\ 40^2 \end{pmatrix}' V_{\hat{\beta}} \begin{pmatrix} 40 \\ 40^2 \end{pmatrix}\right)$$

$$se(\hat{\theta}) = \frac{1}{\sqrt{n}}\sqrt{\hat{V\theta}}$$

$$V_{\hat{\theta}} = \begin{pmatrix} 40 \\ 40^2 \end{pmatrix}' V_{\hat{\beta}} \begin{pmatrix} 40 \\ 40^2 \end{pmatrix}$$

Finally, once $t \to_d \mathcal{N}(0,1)$ is acquired, we would choose some threshold, α, such that if $|t|$ is greater than the $\frac{\alpha}{2}$th quantile of its distribution, then we reject the null hypothesis, that a 40-year-old worker has an expected wage of $40 per hour.