Problem Set #3

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

November 30, 2020

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

Question 1

3.24) The table below displays the results of each of the regressions that must be estimated for this question. The first column displays the estimates for equation (3.50), and the fourth column displays the estimates for the residual approach, using the residuals from the regressions displayed in columns (2) and (3).

VARIABLES	(1) log(wage)	(2) log(wage)	(3) education	(4) $\hat{\varepsilon}_{wage}$
Education	0.144***			
	(0.0116)			
Experience	0.0426***	0.0448***	0.0150	
	(0.0122)	(0.0153)	(0.0643)	
Experience^2	-0.0951***	-0.116***	-0.143	
	(0.0349)	(0.0438)	(0.184)	
$\hat{\varepsilon}_{educ}$				0.144***
				(0.0116)
Constant	0.531***	2.679***	14.89***	-2.33e-09
	(0.190)	(0.0973)	(0.409)	(0.0341)
Observations	267	267	267	267
R-squared	0.389	0.033	0.014	0.369
Sum-of-squared Errors	82.50	130.7	2314	82.50

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) As shown in the table above, $R^2 = 0.389$ and the sum of squared errors is 82.50.

(b) Comparing the coefficients from the "Education" row in column (1) and the $\varepsilon_{\hat{educ}}$ row of column (4) shows that the coefficient on education is the same using either a partition regression or a residual regression. They each equal 0.144.

(c) The bottom three rows of the table above provide summary statistics for each regression. These show that the sum-of-squared errors for the regressions in (a) and (b) are the same, but the R^2 is slightly smaller when the residual regression approach is used than when a partition regression is used. This is to be expected, as using a regression with more independent variables always weakly increases R^2, and the residual regression uses two fewer independent variables than the partition regression.

3.25) Each of the values below is rounded to the nearest thousandth.

(a) $\sum_{i=1}^n \hat{e}_i = 0$

(b) $\sum_{i=1}^n X_{1i}\hat{e}_i = 0$

(c) $\sum_{i=1}^n X_{2i}\hat{e}_i = 0$

(d) $\sum_{i=1}^n X_{1i}^2\hat{e}_i = 133.133$

(e) $\sum_{i=1}^n X_{2i}^2\hat{e}_i = 0$

(f) $\sum_{i=1}^n \hat{Y}_i\hat{e}_i = 0$

(g) $\sum_{i=1}^n \hat{e}_i^2 = 82.505$

Question 2

7.2) To find the limit of $\hat{\beta}$ as $n \to \infty$, we can first rewrite $\hat{\beta}$ in terms of expectation:

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{i=1}^n X_i X_i' + \lambda I_k\right)^{-1} \left(\sum_{i=1}^n X_i Y_i\right) \\
&= \left(\sum_{i=1}^n X_i X_i' + \lambda I_k\right)^{-1} n \left(\frac{1}{n}\sum_{i=1}^n X_i Y_i\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^n X_i X_i' + \frac{1}{n}\lambda I_k\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n X_i Y_i\right)
\end{aligned}$$

And, recognizing $Y_i = X_i\beta + \varepsilon_i$,

$$\begin{aligned}
\hat{\beta} \to_p & \mathbb{E}\left(X_i X_i' + 0 I_k\right)^{-1} \mathbb{E}\left(X_i(X_i\beta + \varepsilon_i)\right) \\
= & \mathbb{E}\left(X_i X_i'\right)^{-1} \left[\mathbb{E}\left(X_i X_i'\beta\right) + \mathbb{E}\left(X_i\varepsilon_i\right)\right)] \\
= & \beta\, \mathbb{E}\left(X_i X_i'\right)^{-1} \mathbb{E}\left(X_i X_i'\right) \\
= & \beta
\end{aligned}$$

7.3) Let $\lambda = cn$ where $c > 0$. Then,

$$\begin{aligned}
\hat{\beta} = & \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i' + \frac{1}{n}(cn) I_k\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i Y_i\right) \\
\hat{\beta} \to_p & \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \mathbb{E}\left(X_i(X_i\beta + \varepsilon_i)\right) \\
= & \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \left[\mathbb{E}\left(X_i X_i'\beta\right) + \mathbb{E}\left(X_i\varepsilon_i\right)\right)] \\
= & \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \beta\, \mathbb{E}\left(X_i X_i'\right) \\
= & \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \beta \left[\mathbb{E}\left(X_i X_i'\right) + cI_k - cI_k\right] \\
= & \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \left[\mathbb{E}\left(X_i X_i'\right) + cI_k\right] \beta - \lambda\beta \\
= & \beta - c\, \mathbb{E}\left(X_i X_i' + cI_k\right)^{-1} \beta
\end{aligned}$$

7.4) (a) $\mathbb{E}(X_1) = \frac{4}{8}(-1) + \frac{4}{8}(1) = 0$

(b) $\mathbb{E}(X_1^2) = \frac{4}{8}(-1)^2 + \frac{4}{8}(1)^2 = 1$

(c) $\mathbb{E}(X_1 X_1) = \frac{3}{8}(1)(1) + \frac{3}{8}(-1)(-1) + \frac{1}{8}(1)(-1) + \frac{1}{8}(-1)(1) = \frac{6}{8} - \frac{2}{8} = \frac{1}{2}$

(d) $\mathbb{E}(e^2) = \left(\frac{3}{4}\right)\left(\frac{5}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{15}{16} + \frac{1}{16} = 1$

(e) $\mathbb{E}(X_1^2 e^2) = \left(\frac{3}{4}\right)\left(\frac{5}{4}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = 1$

(f) $\mathbb{E}(X_1 X_1 e^2) = \left(\frac{3}{4}\right)\left(\frac{5}{4}\right)(1) + \left(\frac{1}{4}\right)\left(\frac{1}{4}\right)(-1) = \frac{15}{16} - \frac{1}{16} = \frac{7}{8}$

Question 3

Note that $\varepsilon^2 \to_p \sigma^2$ and $\hat{sigma}^2 = \hat{\varepsilon}^2$. Then, using the notation from 6.8, where $Z_n \to_p 0 \equiv Z_n = o_p(1)$ and $X_n = O_p(1)$ means that Z_n is bounded in probability about 0,

$$
\begin{aligned}
\sqrt{N}\left(\hat{\sigma}^2 - \sigma^2\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \left(x_i'\beta + \varepsilon_i - x_i'\hat{\beta}\right)^2 - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \left(\varepsilon_i - x_i'(\hat{\beta} - \beta)\right)^2 - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^{n} x_i'\varepsilon_i\right)\sqrt{n}(\hat{\beta} - \beta) + \sqrt{n}(\hat{\beta} - \beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)(\hat{\beta} - \beta)
\end{aligned}
$$

$$
\begin{aligned}
\left(\frac{1}{n}\sum_{i=1}^{n} x_i'\varepsilon_i\right) &\to_p 0 \Rightarrow \left(\frac{1}{n}\sum_{i=1}^{n} x_i'\varepsilon_i\right) = o_p(1) \\
\sqrt{n}(\hat{\beta} - \beta) &\to_d \mathcal{N}(0, V) \Rightarrow \sqrt{n}(\hat{\beta} - \beta) = O_p(1) \\
\hat{\beta} - \beta &\to_p 0 \Rightarrow \hat{\beta} - \beta = o_p(1) \\
\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right) &\to_p \mathbb{E}(x_i x_i') = X'X \\
X'X(\hat{\beta} - \beta) &= X'X\hat{\beta} - X'X\beta = X'X(X'X)^{-1}X'Y - X'X\beta = X'(X\beta + \varepsilon) - X'X\beta = X'\varepsilon = 0 \\
\sqrt{N}\left(\hat{\sigma}^2 - \sigma^2\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - 2o_p(1)O_p(1) + O_p(1)o_p(1) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - 2o_p(1) + o_p(1) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - (o_p(1) + o_p(1)) + o_p(1) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) + o_p(1) \\
&\to_d \mathcal{N}(0, \mathbb{E}(\varepsilon_i^4) - \sigma^4)
\end{aligned}
$$

Question 4

I show below that both estimators, $\hat{\beta}$ and $\widetilde{\beta}$ are consistent for β.

$$\begin{aligned}
\hat{\beta} =& \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \frac{n^{-1}\sum_{i=1}^n X_i(X_i'\beta + e_i)}{n^{-1}\sum_{i=1}^n X_i^2} \\
\to_p & \frac{\mathbb{E}(X_i(X_i'\beta + e_i))}{\mathbb{E}(X_i X_i')} = \frac{\mathbb{E}(X_i X_i'\beta) + \mathbb{E}(X_i' e_i)}{\mathbb{E}(X_i X_i')} \\
=& \beta\frac{\mathbb{E}(X_i X_i')}{\mathbb{E}(X_i X_i')} + \frac{\mathbb{E}(X_i' e_i))}{\mathbb{E}(X_i X_i')} = \beta(1) + 0
\end{aligned}$$

$$\therefore \hat{\beta} \to_p \beta$$

$$\begin{aligned}
\widetilde{\beta} =& \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{X_i} = \frac{1}{n}\sum_{i=1}^n \frac{X_i'\beta + e_i}{X_i} = \frac{1}{n}\sum_{i=1}^n \beta\left(1 + \frac{e_i}{X_i}\right) \\
\to_p & \beta\,\mathbb{E}\left(1 + \frac{e_i}{X_i}\right) = \beta\left(1 + \mathbb{E}\left(\frac{1}{X_i}\mathbb{E}(e_i|X)\right)\right) = \beta(1+0)
\end{aligned}$$

$$\therefore \widetilde{\beta} \to_p \beta$$

Question 5

(a) $\hat{Y_{n+1}} = x'\hat{\beta}$, where, as stated in the question, $X_{n+1} = x$.

(b) Our goal is to calculate $\hat{\sigma}_{n+1}^2 = \hat{\varepsilon}_{n+1}^2$. First, we must find $\hat{\varepsilon}$:

$$\hat{\varepsilon}_{n+1} = Y_{n+1} - \hat{Y}_{n+1} = x'\beta + \varepsilon_{n+1} - x'\hat{\beta} = \varepsilon_{n+1} - x'(\hat{\beta} - \beta)$$

Then, we can calculate:

$$\begin{aligned}
\hat{\sigma}_{n+1}^2 &= \left(\varepsilon_{n+1} - x'(\hat{\beta} - \beta)\right)^2 \\
&= \varepsilon_{n+1}^2 - 2\varepsilon_{n+1}x'(\hat{\beta} - \beta) + x'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'x \\
&= \varepsilon_{n+1}^2 - 2\varepsilon_{n+1}x'\hat{\beta} + 2\varepsilon_{n+1}x'\beta + x'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'x
\end{aligned}$$

Since $\hat{\beta}$ was chosen in a sample that excludes x, it is independent of x and ε_{n+1}. Thus,

$$\mathbb{E}(\varepsilon_{n+1}x'\hat{\beta}) = \mathbb{E}(\varepsilon_{n+1}x')\,\mathbb{E}((\hat{\beta}) = 0$$

Then,

$$\begin{aligned}
\mathbb{E}(\hat{\sigma}_{n+1}^2|x) &= \mathbb{E}(\varepsilon_{n+1}^2|x) + \mathbb{E}(2\varepsilon_{n+1}x'\beta|x) + \mathbb{E}(x'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'x|x) \\
&= \mathbb{E}(\varepsilon_{n+1}^2|x) + 2x'\beta\,\mathbb{E}(\varepsilon_{n+1}|x) + x'\,\mathbb{E}((\hat{\beta} - \beta)(\hat{\beta} - \beta)'|x)x \\
&= \hat{\sigma}^2 + x'\hat{V}_{\hat{\beta}}x
\end{aligned}$$

Thus, $\hat{\sigma}_{n+1}^2 = \hat{\sigma}^2 + x'\hat{V}_{\hat{\beta}}x$ is the estimator.

Question 6

7.13) (a) Since X and Y are scalars, we can propose the estimator $ga\hat{m}ma = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{Y_i}$.

(b) Since $\theta = 1/\gamma$, we should propose $\hat{\theta} = 1/\hat{\gamma}$.

(c) First, we should find the asymptotic distribution of $\hat{\gamma}$:

$$\sqrt{n}(\hat{\gamma} - \gamma) = \mathcal{N}(0, V)$$

Where:

$$\begin{aligned} V &= \sum_{i=1}^{n} Var(X_i/Y_i) = Var\left(\frac{Y_i\gamma + \mu_i}{Y_i}\right) \\ &= Var\left(\frac{Y_i\gamma + \mu_i}{Y_i}\right) = Var\left(\gamma + \frac{\mu_i}{Y_i}\right) \\ &= \frac{Var(\mu_i)}{Var(Y_i)} \end{aligned}$$

Then, using the delta method, we can find the asymptotic distribution of $\hat{\theta}$, where $\theta = f(gamma) = 1/\gamma$:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &\to_d f'(\gamma)\mathcal{N}\left(0, \frac{Var(\mu_i)}{Var(Y_i)}\right) \\ &= \mathcal{N}\left(0, \frac{1}{\gamma^4}\frac{Var(\mu_i)}{Var(Y_i)}\right) \\ &= \mathcal{N}\left(0, \theta^4\frac{Var(\mu_i)}{Var(Y_i)}\right) \end{aligned}$$

(d) Given the asymptotic variance found in (c), we can caluate the standard error by simply taking the square root of the asymptotic variance: $\frac{1}{\sqrt{n}}\theta^2\sqrt{\frac{Var(\mu_i)}{Var(Y_i)}}$

7.14) (a) The approriate estimator is simply $\hat{\theta} = \hat{\beta}_1\hat{\beta}_2$, where $\hat{\beta}_i$ is the OLS estimator for β_i.

(b) Define θ as a function of β, where $\theta = f(\beta) = \beta_1\beta_2\beta$. Then, using the delta method,

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta) &\to_d \mathcal{N}(0, V) \\ V &= \mathbb{E}(X_iX_i')^{-1}\,\mathbb{E}(\varepsilon_i^2X_iX_i')\,\mathbb{E}(X_iX_i')^{-1} \\ \sqrt{n}(\hat{\theta} - \theta) &\to_d f'(\beta)\mathcal{N}(0, V) \\ &= (\beta_1\beta_2)\mathcal{N}(0, V) \\ &= \mathcal{N}\left(0, (\beta_1\beta_2)V(\beta_1\beta_2)'\right) \end{aligned}$$

(c) An asymptotic 95% confidence interval for θ can be calulated by adding and substracting 1.96 times the standard error of $\hat{\theta}$ from $\hat{\theta}$:

$$\left[\hat{\theta} - 1.96\frac{1}{\sqrt{n}}\sqrt{(\beta_1\beta_2)V(\beta_1\beta_2)'}, \hat{\theta} + 1.96\frac{1}{\sqrt{n}}\sqrt{(\beta_1\beta_2)V(\beta_1\beta_2)'}\right]$$

7.15)

$$\begin{aligned}
\widetilde{\beta} =& \frac{\sum_{i=1}^n X_i^3 Y_i}{\sum_{i=1}^n X_i^4} = \frac{n^{-1}\sum_{i=1}^n X_i^3(X_i\beta + e_i)}{n^{-1}\sum_{i=1}^n X_i^4} \\
=& \beta\left(\frac{n^{-1}\sum_{i=1}^n X_i^4}{n^{-1}\sum_{i=1}^n X_i^4}\right) + \left(\frac{n^{-1}\sum_{i=1}^n X_i^3 e_i}{n^{-1}\sum_{i=1}^n X_i^4}\right) \\
\widetilde{\beta} - \beta =& \left(\frac{\frac{1}{n}\sum_{i=1}^n X_i^3 e_i}{\frac{1}{n}\sum_{i=1}^n X_i^4}\right) \\
\sqrt{n}(\widetilde{\beta} - \beta) =& \left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^3 e_i}{\frac{1}{n}\sum_{i=1}^n X_i^4}\right) \to_d N\left(0, \frac{\mathbb{E}(X^6\varepsilon^2)}{\mathbb{E}(X^4)^2}\right)
\end{aligned}$$

Question 7

(a) The variance of $\hat{\theta} = \hat{\beta}_1 - \hat{\beta}_2$ is $Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)$, which we can use to solve for the standard error of $\hat{\theta}$, which we will denote as $s(\hat{\theta})$:

$$\begin{aligned}
Var(\hat{\theta}) &= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2) = s(\hat{\beta}_1)^2 + s(\hat{\beta}_2)^2 - 2\hat{\rho}s(\hat{\beta}_1)s(\hat{\beta}_2) \\
&= (0.07)^2 + (0.07)^2 - \hat{\rho}(0.07)(0.07) = 2(0.07)^2 - \hat{\rho}(0.07)^2 \\
&= (0.07)^2(1 - \hat{\rho}) \\
s(\hat{\theta}) &= \sqrt{Var(\hat{\theta})} = 0.07\sqrt{1 - \hat{\rho}}
\end{aligned}$$

The mean of $\hat{\theta}$ is simply $\hat{\beta}_1 - \hat{\beta}_2 = 0.2$ Then, the 95% confidence interval for θ is

$$\left[0.2 - 1.96\left(0.07\sqrt{1-\hat{\rho}}\right), 0.2 + 1.96\left(0.07\sqrt{1-\hat{\rho}}\right)\right]$$

(b) No. $\hat{\rho}$ relies on the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$, which cannot be deduced from the variance and mean of each coeffient, which is the only information we have.

(c) While we don't know the value of $\hat{\rho}$, we know that it is between 0 and 1. Thus, we can calculate an interval of all possible values of the lower end of the confidence interval for θ. This interval is

$$\left[0.2 - 1.96\left(0.07\sqrt{1-0}\right), 0.2 - 1.96\left(0.07\sqrt{1-1}\right)\right] = [0.062, 0.2]$$

Thus, if we test the difference between $\hat{\beta}_1$ and $\hat{\beta}_2$ using a 5% significance level, we would reject the null hypothesis that $\hat{\beta}_1 = \hat{\beta}_2$.

Question 8

Both $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimators of β, using different samples. Then, for each beta $j \in \{1, 2\}$,

$$\sqrt{n}(\hat{\beta}_j - \beta) \to_d \mathcal{N}\left(0, \sigma_j^2 (X_{ji} X_{ji}')^{-1}\right)$$

Note that σ^2 depends on the subsample because we do not know whether the relationship between X and Y is homoskedastic.

Then, we can solve for the asymptotic distribution of $\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_2)$:

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_2) &= \sqrt{n}(\hat{\beta}_1 - \beta - \hat{\beta}_2 + \beta) = \sqrt{n}(\hat{\beta}_1 - \beta - (\hat{\beta}_2 - \beta)) \\
&= \sqrt{n}(\hat{\beta}_1 - \beta) - \sqrt{n}(\hat{\beta}_2 - \beta)
\end{aligned}$$

If the asymptotic variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ were independent, we could simply add them. However, they are cleary not asymptotic, as they come from the same data-generating process. Thus, we need to estimate the joint variance. We can do so by re-writing the estimation as:

$$Y_i = d_i x_i' \beta + (1 - d_i) x_i' \beta + \varepsilon_i$$

Where $d_i = \begin{cases} 1, & \text{obs in sample 1} \\ 0, & \text{otherwise} \end{cases}$. Assume that observations are randomly assigned between subsamples such that d_i is independent of (y_i, x_i). Recall that $\hat{\beta} - \beta = (X'X)^{-1} X' \varepsilon$. Then, recognizing that $X_i = \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix}$,

$$\begin{aligned}
\hat{\beta} - \beta &= \left(\frac{1}{2n} \sum_{i=1}^{2n} X_i X_i' \right)^{-1} \frac{1}{2n} \sum_{i=1}^{2n} X_i \varepsilon_i \\
&= \left(\frac{1}{2n} \sum_{i=1}^{2n} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix}' \right)^{-1} \frac{1}{2n} \sum_{i=1}^{2n} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix} \varepsilon_i \\
&= \begin{pmatrix} \frac{1}{2n} \sum_{i=1}^{2n} d_i x_i x_i' & \frac{1}{2n} \sum_{i=1}^{2n} d_i (1 - d_i) x_i x_i' \\ \frac{1}{2n} \sum_{i=1}^{2n} (1 - d_i) d_i x_i x_i' & \frac{1}{2n} \sum_{i=1}^{2n} (1 - d_i)(1 - d_i) x_i x_i' \end{pmatrix}^{-1} \frac{1}{2n} \sum_{i=1}^{2n} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix} \varepsilon_i \\
&= \begin{pmatrix} \frac{1}{2n} \sum_{i=1}^{2n} d_i x_i x_i' & 0 \\ 0 & \frac{1}{2n} \sum_{i=1}^{2n} (1 - d_i)^2 x_i x_i' \end{pmatrix}^{-1} \frac{1}{2n} \sum_{i=1}^{2n} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix} \varepsilon_i \\
\sqrt{n}(\hat{\beta} - \beta) &= \begin{pmatrix} \frac{1}{2n} \sum_{i=1}^{2n} d_i x_i x_i' & 0 \\ 0 & \frac{1}{2n} \sum_{i=1}^{2n} (1 - d_i)^2 x_i x_i' \end{pmatrix}^{-1} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2n}} \sum_{i=1}^{2n} \begin{pmatrix} d_i x_i \\ (1 - d_i) x_i \end{pmatrix} \varepsilon_i
\end{aligned}$$

Recognizing that $\mathbb{E}(d_i) = \frac{1}{2}$,

$$\begin{pmatrix} \frac{1}{2n}\sum_{i=1}^{2n} d_i x_i x_i' & 0 \\ 0 & \frac{1}{2n}\sum_{i=1}^{2n}(1-d_i)^2 x_i x_i' \end{pmatrix}^{-1} \to_p \begin{pmatrix} \frac{1}{2}\,\mathbb{E}(x_i x_i') & 0 \\ 0 & \frac{1}{2}\,\mathbb{E}(x_i x_i') \end{pmatrix}^{-1}$$

$$\frac{1}{\sqrt{2n}}\sum_{i=1}^{2n}\begin{pmatrix} d_i x_i \\ (1-d_i)x_i \end{pmatrix}\varepsilon_i \to_d \mathcal{N}\left(0, \begin{pmatrix} \frac{1}{2}\,\mathbb{E}(\varepsilon^2 x_i x_i') & 0 \\ 0 & \frac{1}{2}\,\mathbb{E}(\varepsilon^2 x_i x_i') \end{pmatrix}\right)$$

Thus,

$$\sqrt{n}(\hat{\beta}-\beta) \to_d \mathcal{N}\left(0, \begin{pmatrix} 2\,\mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(\varepsilon^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1} & 0 \\ 0 & 2\,\mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(\varepsilon^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1} \end{pmatrix}\right)$$

Where the $1/\sqrt{2}$ gets squared when it enters the asymptotic variance, then cancelled out by the 2 that comes from $\left(\frac{1}{2}\,\mathbb{E}(\varepsilon^2 x_i x_i')\right)^{-1}$, which gets squared as it enters the asymptotic variance. This can be simplified to:

$$\sqrt{n}(\hat{\beta}-\beta) \to_d \mathcal{N}\left(0, 2\,\mathbb{E}(x_i x_i')^{-1}\,\mathbb{E}(\varepsilon^2 x_i x_i')\,\mathbb{E}(x_i x_i')^{-1} I_{2n}\right)$$

Question 9

(a) Intuitively, $\hat{\beta}$ is unbiased, since $\mathbb{E}(\mathbb{1}\{x_i \in \{1,2\}\}) = \mathbb{E}(x_i)$. To show that this is the case,

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} w_i y_i \mathbb{1}\{x_i \in \{1,2\}\} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} w_i (w_i'\beta + \varepsilon_i) \mathbb{1}\{x_i \in \{1,2\}\} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} (w_i w_i'\beta + w_i\varepsilon_i) \mathbb{1}\{x_i \in \{1,2\}\} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i'\beta \mathbb{1}\{x_i \in \{1,2\}\} + \frac{1}{n}\sum_{i=1}^{n} w_i\varepsilon_i \mathbb{1}\{x_i \in \{1,2\}\}\right) \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\}\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} w_i\varepsilon_i \mathbb{1}\{x_i \in \{1,2\}\} \\
&\to_p \beta + \mathbb{E}(w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\})^{-1}\,\mathbb{E}(w_i\varepsilon_i \mathbb{1}\{x_i \in \{1,2\}\}) \\
&= \beta + \mathbb{E}(w_i w_i' \mathbb{1}\{x_i \in \{1,2\}\})^{-1}\,\mathbb{E}\left(w_i \mathbb{1}\{x_i \in \{1,2\}\}\,\mathbb{E}(\varepsilon_i|w_i)\right) \\
&= \beta
\end{aligned}$$

Thus, $\hat{\beta} \to_p \beta$.

(b) Note that the cancellation of the second term in the second to last line of the proof in (a) requires $\mathbb{E}(\varepsilon_i|w_i) = 0$. This does not hold with (A1'), so that assumption is not strong enough to conclude that $\hat{\beta} \to_p \beta$.

(c)

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} w_i w_i' \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\}$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\} \to_d \mathcal{N}\left(0, Var\left(w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right)\right)$$

$$\begin{aligned}
Var\left(w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right) &= Var\left(Var\left(w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\}|w_i\right)\right) \\
&= Var\left(w_i \mathbb{1}\left\{x_i \in \{1,2\}\right\} Var\left(\varepsilon_i|w_i\right)\right) \\
&= \sigma^2\, \mathbb{E}\left(w_i w_i' \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right) \\
\mathbb{E}\left(w_i w_i' \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right) &= \mathbb{E}\left[\begin{pmatrix} 1 \\ x_i \end{pmatrix}\begin{pmatrix} 1 \\ x_i \end{pmatrix}' \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right] \\
&= \mathbb{E}\left[\begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right] \\
&= \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{5}{4} \end{pmatrix} \\
\sqrt{n}(\hat{\beta} - \beta) &\to_d \mathbb{E}\left(w_i w_i' \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right)^{-1} \mathcal{N}(0, Var\left(w_i \varepsilon_i \mathbb{1}\left\{x_i \in \{1,2\}\right\}\right)) \\
&= \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}^{-1} \mathcal{N}\left(0, \sigma^2 \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}\right) \\
&= \mathcal{N}\left(0, \sigma^2 \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}^{-1}\right) \\
&= \mathcal{N}\left(0, \sigma^2 \begin{pmatrix} 20 & -12 \\ -12 & 8 \end{pmatrix}\right)
\end{aligned}$$

(d)

(e)

(f)