Problem Set #3 (2^{nd} Half) Solutions

Economics 709 **Fall 2020**

3.24 (a) With the restricted sample, we get the same estimates for the log wage regression as in equation (3.50):

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

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

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

Table 2: OLS Coefficient Estimates for Education on Wage (Using Residual Regression Approach): All Single Asian Males with Less than 45 Years of Experience ($n = 267$)[1]

Variables	First Step Residual
Coefficient estimates	
Second Step Residual	0.144
R^2	0.369
SSE (Sum of Squared Errors)	82.505
observations	267

[1] 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.

The residuals from this residual regression, $\hat{\hat{e}}$, should be same as those from the original regression $\hat{e}$ because

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

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

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

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

since $\hat{\hat{e}} = \hat{e}$, $\tilde{e} = M_{X_2}y$, and $M_{X_2}M_1 = M_{X_2}$. By similar reasoning in Exercise 3.16, $y'M_1y - y'M_{X_2}y = y'\left(M_1 - M_{X_2}\right)y \geq 0$ since $M_1 - M_{X_2}$ is idempotent. Therefore, $y'M_1y \geq y'M_{X_2}y$, and $R^2_{\text{original}} \geq R^2_{\text{residual}}$.

3.25 (a) $\sum_{i=1}^{n}\hat{e}_i \approx 0$.
(b) $\sum_{i=1}^{n}x_{1i}\hat{e}_i \approx 0$.
(c) $\sum_{i=1}^{n}x_{2i}\hat{e}_i \approx 0$.
(d) $\sum_{i=1}^{n}x_{1i}^2\hat{e}_i = x_1^{2\prime}\hat{e} = 133.133$.
(e) $\sum_{i=1}^{n}x_{2i}^2\hat{e}_i \approx 0$.
(f) $\sum_{i=1}^{n}\hat{y}_i\hat{e}_i \approx 0$.
(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. Note that numerical calculations may not return exact zeros.

7.2

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{i=1}^{n}X_iX_i' + \lambda I_k\right)^{-1}\left(\sum_{i=1}^{n}X_iY_i\right) = \left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i' + \frac{1}{n}\lambda I_k\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i(X_i'\beta + \varepsilon_i)\right) \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i' + \frac{1}{n}\lambda I_k\right)^{-1}\left(\left(\frac{1}{n}\sum_{i=1}^{n}X_iX_i'\right)\beta + \left(\frac{1}{n}\sum_{i=1}^{n}X_i\varepsilon_i\right)\right) \\
&\xrightarrow{p} \left(E(X_iX_i') + 0\right)^{-1}\left(E(X_iX_i')\beta + E(X_i\varepsilon_i)\right) = \left(E(X_iX_i')\right)^{-1}\left(E(X_iX_i')\beta + 0\right) = \beta
\end{aligned}$$

using WLLN, $E(X_i\varepsilon_i) = 0$, and $\frac{1}{n}\lambda I_k \longrightarrow 0$. So $\hat{\beta}$ is consistent for β.

7.3

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

So $\hat{\beta}$ is not consistent for β.

7.4 Note $\mathbf{P}(X_{1i} = 1) = \mathbf{P}(X_{1i} = -1) = \mathbf{P}(X_{2i} = 1) = \mathbf{P}(X_{2i} = -1) = \frac{1}{2}$

(a) $E(X_{1i}) = 1 \cdot \mathbf{P}(X_{1i} = 1) + (-1)\mathbf{P}(X_{1i} = -1) = \frac{1}{2} - \frac{1}{2} = 0$

(b) $E(X_{1i}^2) = 1^2 \cdot \mathbf{P}(X_{1i} = 1) + (-1)^2\mathbf{P}(X_{1i} = -1) = \frac{1}{2} + \frac{1}{2} = 1$

(c) $E(X_{1i}X_{2i}) = (1 \cdot 1)\mathbf{P}(X_{1i} = 1, X_{2i} = 1) + (-1 \cdot 1)\mathbf{P}(X_{1i} = -1, X_{2i} = 1)$
$+ (1 \cdot -1)\mathbf{P}(X_{1i} = 1, X_{2i} = -1) + (-1 \cdot -1)\mathbf{P}(X_{1i} = -1, X_{2i} = -1) = \frac{3}{8} - \frac{1}{8} - \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$

(d) $E(e_i^2) = E(e_i^2 | X_{1i} = X_{2i})\mathbf{P}(X_{1i} = X_{2i}) + E(e_i^2 | X_{1i} \neq X_{2i})\mathbf{P}(X_{1i} \neq X_{2i})$
$= \frac{5}{4}\left(\frac{3}{8} + \frac{3}{8}\right) + \frac{1}{4}\left(\frac{1}{8} + \frac{1}{8}\right) = 1$

(e) $E(X_{1i}^2 e_i^2) = E(X_{1i}^2 e_i^2 | X_{1i} = X_{2i})\mathbf{P}(X_{1i} = X_{2i}) + E(X_{1i}^2 e_i^2 | X_{1i} \neq X_{2i})\mathbf{P}(X_{1i} \neq X_{2i})$
$= 1 \cdot \frac{5}{4}\left(\frac{3}{8} + \frac{3}{8}\right) + 1 \cdot \frac{1}{4}\left(\frac{1}{8} + \frac{1}{8}\right) = 1$

(f) $E(X_{1i}X_{2i}e_i^2) = E(X_{1i}X_{2i}e_i^2 | X_{1i} = X_{2i})\mathbf{P}(X_{1i} = X_{2i}) + E(X_{1i}X_{2i}e_i^2 | X_{1i} \neq X_{2i})\mathbf{P}(X_{1i} \neq X_{2i})$
$= 1 \cdot \frac{5}{4}\left(\frac{3}{8} + \frac{3}{8}\right) + (-1) \cdot \frac{1}{4}\left(\frac{1}{8} + \frac{1}{8}\right) = \frac{7}{8}$

7.8 Using the $o_p(1)$ and $O_p(1)$ notation from the book,

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}^2 - \sigma^2) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 - \sigma^2\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - 2\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i x_i{}'\right)\sqrt{n}(\hat{\beta} - \beta) + \sqrt{n}(\hat{\beta} - \beta)'\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i{}'\right)(\hat{\beta} - \beta) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) - 2o_p(1)O_p(1) + O_p(1)O_p(1)o_p(1) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \sigma^2\right) + o_p(1) \xrightarrow{d} N(0, Var(\varepsilon_i^2)) \sim N(0, E(\varepsilon_i^4) - (\sigma^2)^2)
\end{aligned}$$

7.9 (a) We know OLS ($\hat{\beta}$) is consistent.

$$\tilde{\beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i\beta + \varepsilon_i}{X_i} = \beta + \frac{1}{n}\sum_{i=1}^{n} \frac{\varepsilon_i}{X_i} \xrightarrow{p} \beta + E\left(\frac{\varepsilon_i}{X_i}\right) = \beta + E\left(\frac{E(\varepsilon_i | X_i)}{X_i}\right) = \beta$$

so $\tilde{\beta}$ is consistent.

7.10 (a) The out-of-sample forecast is $\hat{y}_{n+1} = \mathbf{X}'_{n+1}\hat{\beta} = \mathbf{x}'\hat{\beta}$.

(b) The variance of the forecast is $Var(\mathbf{x}'\hat{\beta}|\mathbf{X}_{n+1} = \mathbf{x}) = \mathbf{x}'V_{\hat{\beta}}\mathbf{x}$. So an estimator is $\mathbf{x}'\widehat{V}_{\hat{\beta}}\mathbf{x}$.

7.13 (a) Try OLS X_i on Y_i: $\hat{\gamma} = \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n Y_i^2}$

(b) $\hat{\theta} = 1/\hat{\gamma}$

(c) $\sqrt{n}(\hat{\gamma} - \gamma) = \sqrt{n}\frac{\sum_{i=1}^n Y_i u_i}{\sum_{i=1}^n Y_i^2} \xrightarrow{d} N\left(0, \frac{E(Y_i^2 u_i^2)}{E(Y_i^2)^2}\right)$
Then, by the Delta Method,
$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\left(\frac{1}{\hat{\gamma}} - \frac{1}{\gamma}\right) \approx -\gamma^{-2}\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} \gamma^{-2}N\left(0, \frac{E(Y_i^2 u_i^2)}{E(Y_i^2)^2}\right) \sim N\left(0, \frac{E(Y_i^2 u_i^2)}{\gamma^4 E(Y_i^2)^2}\right)$

(d) The asymptotic standard error is $\sqrt{\frac{E(Y_i^2 u_i^2)}{\gamma^4 E(Y_i^2)^2}}$

7.14 (a) $\hat{\theta} = \hat{\beta}_1\hat{\beta}_2$, where $(\hat{\beta}_1, \hat{\beta}_2)'$ is the OLS estimator.

(b) Given the asymptotic distribution for OLS, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_\beta)$, where $V_\beta = E(X_i X_i')^{-1}E(\varepsilon_i^2 X_i X_i')E(X_i X_i')^{-1}$. Then, by the Delta Method,
$\sqrt{n}(\hat{\theta} - \theta) \approx [\beta_2 \;\; \beta_1]\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, [\beta_2 \;\; \beta_1]V_\beta[\beta_2 \;\; \beta_1]')$

(c) Let $\hat{t} = \sqrt{[\hat{\beta}_2 \;\; \hat{\beta}_1]\hat{V}_\beta[\hat{\beta}_2 \;\; \hat{\beta}_1]'}$. Then, the 95% CI is:
$[\hat{\theta} - 1.96\hat{t}/\sqrt{n}, \hat{\theta} + 1.96\hat{t}/\sqrt{n}]$

7.15 $\hat{\beta} = \frac{\sum_{i=1}^n X_i^3 Y_i}{\sum_{i=1}^n X_i^4} = \beta + \frac{\sum_{i=1}^n X_i^3 \varepsilon_i}{\sum_{i=1}^n X_i^4}$. Then,
$\sqrt{n}(\hat{\beta} - \beta) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^3 \varepsilon_i}{\frac{1}{n}\sum_{i=1}^n X_i^4} \xrightarrow{d} \frac{1}{E(X_i^4)}N(0, E(\varepsilon_i^2 X_i^6)) \sim N\left(0, \frac{E(\varepsilon_i^2 X_i^6)}{E(X_i^4)^2}\right)$
since $E(X_i^3\varepsilon_i) = E[X_i^3 E(\varepsilon_i|X_i)] = 0$ so the CLT yields $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^3\varepsilon_i \xrightarrow{d} N(0, E(\varepsilon_i^2 X_i^6))$, and Slutsky's Lemma gives the result.

7.17 (a) $Var(\hat{\theta}) = Var(\hat{\beta}_1 - \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2) = s(\hat{\beta}_1)^2 + s(\hat{\beta}_2)^2 - 2\hat{\rho}s(\hat{\beta}_1)s(\hat{\beta}_2) = 2(0.07)^2 - 2\hat{\rho}(0.07)^2 = 2(0.07)^2(1 - \hat{\rho})$. Then, the 95% CI is:
$[\hat{\theta} - 1.96\sqrt{Var(\hat{\theta})}, \hat{\theta} + 1.96\sqrt{Var(\hat{\theta})}] = [0.2 - 1.96\sqrt{2(0.07)^2(1-\hat{\rho})}, 0.2 + 1.96\sqrt{2(0.07)^2(1-\hat{\rho})}]$

(b) No, $\hat{\rho} = Corr(\hat{\beta}_1, \hat{\beta}_2) = \frac{Cov(\hat{\beta}_1, \hat{\beta}_2)}{s(\hat{\beta}_1)s(\hat{\beta}_2)}$. Only the denominator values are given, but not the numerator.

(c) One way to assess the claim is to check if zero is contained in the 95% CI. But, since $\hat{\rho}$ cannot be calculated, the exact 95% CI is not known from the given information. We do know that $-1 \leq \hat{\rho} \leq 1$. As $\hat{\rho}$ varies over this range, the 95% CI is largest when $\hat{\rho} = -1$ and smallest when $\hat{\rho} = 1$. The largest 95% CI is $[0.2 - 1.96(0.14), 0.2 + 1.96(0.14)]$ which does contain zero, and the smallest CI is the single point $\{0.2\}$, which does not contain zero. So it is not clear if the author's claim is correct.

7.19 Let

$$d_i = \begin{cases} 1 & \text{if } i \text{ is chosen to be in the first split} \\ 0 & \text{if } i \text{ is chosen to be in the second split} \end{cases}$$

With random splitting, we'll take d_i to be independent of (Y_i, X_i). Now we can rewrite our regression equation:

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

$$\sqrt{n}\left[\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}-\begin{pmatrix}\beta\\ \beta\end{pmatrix}\right]=\left[\frac{1}{2n}\sum_i\begin{pmatrix}d_iX_i\\ (1-d_i)X_i\end{pmatrix}\begin{pmatrix}d_iX_i\\ (1-d_i)X_i\end{pmatrix}'\right]^{-1}\frac{1}{2\sqrt{n}}\sum_i\begin{pmatrix}d_iX_i\varepsilon_i\\ (1-d_i)X_i\varepsilon_i\end{pmatrix}$$

$$=\left[\begin{array}{cc}\frac{1}{2n}\sum_i d_i^2X_iX_i' & \frac{1}{2n}\sum_i d_i(1-d_i)X_iX_i'\\ \frac{1}{2n}\sum_i d_i(1-d_i)X_iX_i' & \frac{1}{2n}\sum_i(1-d_i)^2X_iX_i'\end{array}\right]^{-1}\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2n}}\sum_i\begin{pmatrix}d_iX_i\varepsilon_i\\ (1-d_i)X_i\varepsilon_i\end{pmatrix}$$

$$=\left[\begin{array}{cc}\frac{1}{2n}\sum_i d_iX_iX_i' & 0\\ 0 & \frac{1}{2n}\sum_i(1-d_i)X_iX_i'\end{array}\right]^{-1}\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2n}}\sum_i\begin{pmatrix}d_iX_i\varepsilon_i\\ (1-d_i)X_i\varepsilon_i\end{pmatrix}$$

Using $E(d_i)=1/2$,

$$\left[\begin{array}{cc}\frac{1}{2n}\sum_i d_iX_iX_i' & 0\\ 0 & \frac{1}{2n}\sum_i(1-d_i)X_iX_i'\end{array}\right]\xrightarrow{p}\left[\begin{array}{cc}\frac{1}{2}E(X_iX_i') & 0\\ 0 & \frac{1}{2}E(X_iX_i')\end{array}\right]$$

and

$$\frac{1}{\sqrt{2n}}\sum_i\begin{pmatrix}d_iX_i\varepsilon_i\\ (1-d_i)X_i\varepsilon_i\end{pmatrix}\xrightarrow{d}N\left(\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}\frac{1}{2}E(\varepsilon_i^2X_iX_i') & 0\\ 0 & \frac{1}{2}E(\varepsilon_i^2X_iX_i')\end{pmatrix}\right)$$

Then, Slutsky's Lemma yields

$$\sqrt{n}\left[\begin{pmatrix}\hat\beta_1\\ \hat\beta_2\end{pmatrix}-\begin{pmatrix}\beta\\ \beta\end{pmatrix}\right]\xrightarrow{d}N\left(\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}E(X_iX_i')^{-1}E(\varepsilon_i^2X_iX_i')E(X_iX_i')^{-1} & 0\\ 0 & E(X_iX_i')^{-1}E(\varepsilon_i^2X_iX_i')E(X_iX_i')^{-1}\end{pmatrix}\right)$$

Hence,

$$\sqrt{n}(\hat\beta_1-\hat\beta_2)=\sqrt{n}[(\hat\beta_1-\beta)-(\hat\beta_2-\beta)]\xrightarrow{d}N(0,2E(X_iX_i')^{-1}E(\varepsilon_i^2X_iX_i')E(X_iX_i')^{-1})$$

This result is correct, but actually the argument above has one aspect that is problematic. Because exactly half the observations are in each split, the d_i are not i.i.d. So, the Khintchine WLLN and Lindeberg-Levy CLT for i.i.d. data do not apply. Fortunately, there is a straightforward way to "fix" up the argument. We can derive the convergence in probability and convergence in distribution results above conditional on $D=(d_1,\dots,d_{2n})$ and apply forms of the WLLN and CLT that impose (conditional) independence but allow the distributions to change with n (e.g. Markov LLN, Liapunov CLT). Application of these results will yield the same asymptotic distribution as above. Since the conditional distribution of $\sqrt{n}(\hat\beta_1-\hat\beta_2)$ has a limit that does not depend on the conditioning variables, the unconditional limiting distribution is exactly as given above.

9. (a) Yes. Let $d_i=\mathbf{1}\{x_i\in\{1,2\}\}$. By (A0) and WLLN,
   - $E(d_i)=\mathbf{P}(x_i\in\{1,2\})=\frac{1}{2},\ \ E(x_id_i)=\frac{1+2}{4}=\frac{3}{4},\ \ E(x_i^2d_i)=\frac{1+4}{4}=\frac{5}{4}$
   - $\frac{1}{n}\sum_{i=1}^n w_i w_i'd_i\xrightarrow{p}E(w_i w_i'd_i)=\begin{pmatrix}E(d_i) & E(x_id_i)\\ E(x_id_i) & E(x_i^2d_i)\end{pmatrix}=\begin{pmatrix}1/2 & 3/4\\ 3/4 & 5/4\end{pmatrix}$ which is finite and nonsingular
   - $\frac{1}{n}\sum_{i=1}^n w_i\varepsilon_id_i\xrightarrow{p}E(w_i\varepsilon_id_i)=E[w_iE(\varepsilon_i|w_i)d_i]=E[w_i\cdot 0\cdot d_i]=0$
   - $\hat\beta=\left[\frac{1}{n}\sum_{i=1}^n w_i w_i'd_i\right]^{-1}\frac{1}{n}\sum_{i=1}^n w_i y_id_i=\left[\frac{1}{n}\sum_{i=1}^n w_i w_i'd_i\right]^{-1}\frac{1}{n}\sum_{i=1}^n w_i(w_i'\beta+\varepsilon_i)d_i=\beta+\left[\frac{1}{n}\sum_{i=1}^n w_i w_i'd_i\right]^{-1}\frac{1}{n}\sum_{i=1}^n w_i\varepsilon_id_i\xrightarrow{p}\beta+[E(w_i w_i'd_i)]^{-1}E(w_i\varepsilon_id_i)=\beta$

   (b) No, not in general. (A1′): $E(w_i\varepsilon_i)=0$ does *not* imply $E(w_i\varepsilon_id_i)=0$, so consistentency may fail.

(c) • $Var(w_i\varepsilon_i d_i) = E(w_i w_i' \varepsilon_i^2 d_i) = E[w_i w_i' E(\varepsilon_i^2|w_i)d_i] = \sigma^2 E[w_i w_i' d_i]$

• From (a), $\sqrt{n}(\hat{\beta} - \beta) = \left[\frac{1}{n}\sum_{i=1}^n w_i\, w_i' d_i\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n w_i \varepsilon_i d_i \xrightarrow{d} E[w_i w_i' d_i]^{-1} N(0, \sigma^2 E[w_i w_i' d_i]) \sim N(0, \sigma^2 E[w_i w_i' d_i]^{-1}) \sim N\left(0, \sigma^2 \begin{pmatrix} 1/2 & 3/4 \\ 3/4 & 5/4 \end{pmatrix}^{-1}\right) \sim N\left(0, \sigma^2 \begin{pmatrix} 20 & -12 \\ -12 & 8 \end{pmatrix}\right)$

(d) Under (A0) and (A1), $\hat{\beta}_2$ and $\hat{\hat{\beta}}_2$ are both unbiased and consistent estimators for γ. Let's compare their asymptotic variances. From (c), note that $\sqrt{n}(\hat{\beta}_2 - \beta_2) \xrightarrow{d} N(0, \sigma^2 [E(w_i w_i' d_i)^{-1}]_{2,2})$. And, $[E(w_i w_i' d_i)^{-1}]_{2,2} = \frac{E(d_i)}{E(x_i^2 d_i)E(d_i) - [E(x_i d_i)]^2} = \frac{Pr(d_i=1)}{Var(x_i|d_i=1)\mathbf{P}(d_i=1)^2} = \frac{1}{Var(x_i|d_i=1)\mathbf{P}(d_i=1)}$.
Similarly, we can show that $\sqrt{n}(\hat{\hat{\beta}}_2 - \beta_2) \xrightarrow{d} N(0, \sigma^2 [E(w_i w_i'(1-d_i))^{-1}]_{2,2})$ and $[E(w_i w_i'(1-d_i))^{-1}]_{2,2} = \frac{1}{Var(x_i|d_i=0)\mathbf{P}(d_i=0)}$. $Var(x_i|d_i=1) > Var(x_i|d_i=0)$ ($\frac{1}{4} > \frac{1}{36}$), so AVAR($\sqrt{n}(\hat{\beta}_2 - \beta_2)) = \frac{\sigma^2}{Var(x_i|d_i=1)\mathbf{P}(d_i=1)} < \frac{\sigma^2}{Var(x_i|d_i=0)\mathbf{P}(d_i=0)} =$ AVAR($\sqrt{n}(\hat{\hat{\beta}}_2 - \beta_2)$) ($8\sigma^2 < 72\sigma^2$).
Explanation: The x distribution with larger variance provides more precise estimates of the slope (all else equal).

(e) $E(x_i y_i d_i) = E[x_i(1 + \gamma x_i + \varepsilon)d_i] = E(x_i d_i) + \gamma E(x_i^2 d_i) + E(x_i \varepsilon_i d_i)$.
$\hat{\alpha} = \left[\frac{1}{n}\sum_{i=1}^n x_i^2 d_i\right]^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i d_i \xrightarrow{p} E(x_i^2 d_i)^{-1} E(x_i y_i d_i) = E(x_i^2 d_i)^{-1}[E(x_i d_i) + \gamma E(x_i^2 d_i)] = \gamma + E(x_i^2 d_i)^{-1} E(x_i d_i) = \gamma + \frac{4}{5} \cdot \frac{3}{4} = \gamma + \frac{3}{5}$

(f) • $\alpha = \gamma + \frac{3}{5}$

• $E(x_i^3 d_i) = \frac{1+8}{4} = \frac{9}{4}$; $E(x_i^4 d_i) = \frac{1+16}{4} = \frac{17}{4}$

• $E[(x_i - \frac{3}{5}x_i^2 + x_i\varepsilon_i)d_i] = 0$ from part (e)

• $Var[(x_i - \frac{3}{5}x_i^2 + x_i\varepsilon_i)d_i] = E[(x_i - \frac{3}{5}x_i^2 + x_i\varepsilon_i)^2 d_i] = E(x_i^2 d_i) + \frac{9}{25}E(x_i^4 d_i) + E(x_i^2 \varepsilon_i^2 d_i)$
$-2 \cdot \frac{3}{5}E(x_i^3 d_i) + 2E(x_i^2 \varepsilon_i d_i) - 2 \cdot \frac{3}{5}E(x_i^3 \varepsilon_i d_i) = E(x_i^2 d_i) - \frac{6}{5}E(x_i^3 d_i) + \frac{9}{25}E(x_i^4 d_i) + E(x_i^2 E(\varepsilon_i^2|w_i)d_i)$
$+2E(x_i^2 E(\varepsilon_i|w_i)d_i) - 2 \cdot \frac{3}{5}E(x_i^3 E(\varepsilon_i|w_i)d_i) = \frac{5}{4} - \frac{6}{5} \cdot \frac{9}{4} + \frac{9}{25} \cdot \frac{17}{4} + \sigma^2 \frac{5}{4} + 0$

• $\sqrt{n}(\hat{\alpha} - \alpha) = \left[\frac{1}{n}\sum_{i=1}^n x_i^2 d_i\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i y_i d_i - x_i^2 d_i \alpha) = \left[\frac{1}{n}\sum_{i=1}^n x_i^2 d_i\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i + x_i^2(\gamma - \alpha) + x_i\varepsilon_i)d_i = \left[\frac{1}{n}\sum_{i=1}^n x_i^2 d_i\right]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i - \frac{3}{5}x_i^2 + x_i\varepsilon_i)d_i$
$\xrightarrow{d} N\left(0, \left(\frac{4}{5}\right)^2 \left(\frac{5}{4} - \frac{6}{5} \cdot \frac{9}{4} + \frac{9}{25} \cdot \frac{17}{4} + \sigma^2 \frac{5}{4}\right)\right) \sim N\left(0, \frac{16}{25}\left(\frac{2}{25} + \frac{5}{4}\sigma^2\right)\right)$.