P218 Problem Set 1

Will Hotten

Question 1

1.a.

$$\frac{\partial Y}{\partial X} = -2X + 2 + \varepsilon$$

$$= \begin{cases} -2X + 3 & \text{if } \varepsilon = 1 \\ -2X + 1 & \text{if } \varepsilon = -1 \end{cases}$$

The causal effect of a marginal change in X on Y is always greater for farms with very good soil compared to those with very bad soil.

1.b.

$$ACE(X, \varepsilon) = E[C(X, \varepsilon)|X]$$

$$= \int_{-1}^{1} \frac{\partial}{\partial X} Y(X, \varepsilon) \cdot f_{\varepsilon|X}(\varepsilon \mid X)\, d\varepsilon$$

For this we need to calculate the density of X conditional on ε:

$$f_{\varepsilon|X}(\varepsilon \mid X) = \frac{f_{X\varepsilon}(x, \varepsilon)}{f_X(x)}$$

$$= \left(\frac{1 + x - \varepsilon}{8}\right) \cdot \left(\frac{1}{f_X(x)}\right)$$

To calculate the marginal density of X:

$$f_X(x) = \int_{-1}^{1} \left(\frac{1 + x - \varepsilon}{8}\right) d\varepsilon$$

$$= \left[\frac{-\varepsilon(1 + x)}{8} + \frac{-\varepsilon^2}{16}\right]_{-1}^{1}$$

$$= -\left[\frac{1 + x}{8} + \frac{1}{16}\right] + \left[-\frac{1 + x}{8} + \frac{1}{16}\right]$$
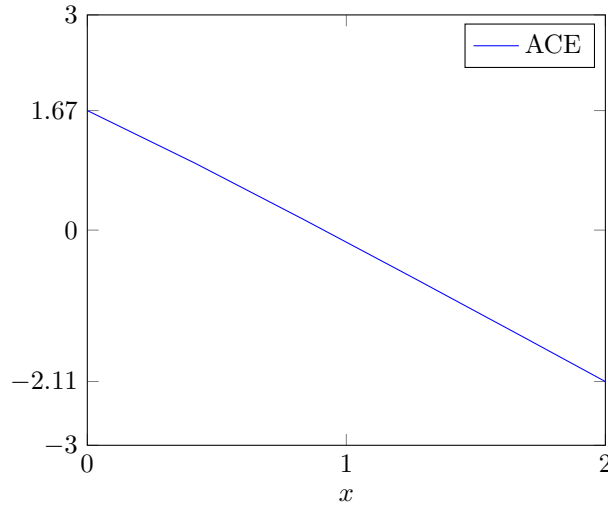
$$= \frac{1 + x}{4}$$

Therefore, the conditional density can be rewritten:

$$f_{\varepsilon|X}(\varepsilon \mid X) = \left(\frac{1+x-\varepsilon}{8}\right) \cdot \left(\frac{4}{1+x}\right)$$
$$= \frac{1}{2} \cdot \left(\frac{1+x-\varepsilon}{1+x}\right)$$

Plugging this into the formula for ACE(X,ε):

$$ACE(X, \varepsilon) = \int_{-1}^{1} \frac{\partial}{\partial X} Y(X, \varepsilon) \cdot \frac{1}{2} \cdot \left(\frac{1+x-\varepsilon}{1+x}\right) d\varepsilon$$
$$= \int_{-1}^{1} \frac{1}{2} \cdot (-2x + 2 + \varepsilon) \cdot \left(\frac{1+x-\varepsilon}{1+x}\right) d\varepsilon$$
$$= \left[\frac{-\varepsilon \cdot \left(\varepsilon \cdot (2\varepsilon - 9x + 3) + 12x^2 - 12\right)}{12(x+1)}\right]_{-1}^{1}$$
$$= \left[\frac{-12x^2 + 9x + 13}{12(x+1)}\right] - \left[\frac{12x^2 + 9x + 7}{12(x+1)}\right]$$
$$= \frac{5 - 6x^2}{3x + 3}$$

Plotting ACE(X,ε):

The average causal effect of X on Y is decreasing in X.

1.c.

E(Y):

$$E(Y) = E(-X^2 + 2X + \varepsilon X)$$
$$= -E(X^2) + 2E(X) + E(\varepsilon X)$$

Now, $E(X)$:

$$E(X) = \int_0^2 x f_x(x)\, dx$$
$$= \int_0^2 x \frac{1+x}{4}$$
$$= \frac{1}{4}\left[\frac{x^2}{2} + \frac{x^3}{3}\right]_0^2$$
$$= \frac{7}{6}$$

$E(X^2)$:

$$E(X^2) = \int_0^2 x^2 f_x(x)\, dx$$
$$= \int_0^4 x^2 \frac{1+x}{4}$$
$$= \frac{1}{4}\left[\frac{x^3}{3} + \frac{x^4}{4}\right]_0^4$$
$$= \frac{5}{3}$$

$E(\varepsilon X)$:

$$E(\varepsilon X) = \int_{-1}^1 \int_0^2 x\varepsilon f_{x\varepsilon}(x, \varepsilon)\, dx\, d\varepsilon$$
$$= \int_{-1}^1 \left[\frac{\varepsilon x^2 \cdot (2x - 3\varepsilon + 3)}{48}\right]_0^2 d\varepsilon$$
$$= \int_{-1}^1 -\frac{\varepsilon \cdot (6\varepsilon - 14)}{24}\, d\varepsilon$$
$$= \left[-\frac{\varepsilon^2 \cdot (2\varepsilon - 7)}{24}\right]_{-1}^1$$
$$= -\frac{1}{6}$$

3

Putting together:

$$E(Y) = -\frac{5}{3} + 2\left(\frac{7}{6}\right) + -\frac{1}{6}$$
$$= \frac{1}{6} \cdot (-10 + 14 - 1)$$
$$= \frac{1}{2}$$

E(Y | X):

$$E(Y \mid X) = E(-X^2 + 2X + \varepsilon X \mid X)$$
$$= -X^2 + 2X + XE(\varepsilon \mid X)$$

From earlier:

$$f_{\varepsilon\mid X}(\varepsilon \mid X) = \frac{1}{2} \cdot \left(\frac{1 + x - \varepsilon}{1 + x}\right)$$

Therefore, $E(\varepsilon \mid X)$:

$$E(\varepsilon \mid X) = \int_{-1}^{1} \varepsilon f_{\varepsilon\mid X}(\varepsilon \mid X)\, d\varepsilon$$
$$= \int_{-1}^{1} \varepsilon \frac{1 + x - \varepsilon}{2(1 + x)}\, d\varepsilon$$
$$= \left[-\frac{\varepsilon^2 \cdot (2\varepsilon - 3x - 3)}{12\,(x + 1)}\right]_0^2$$
$$= -\frac{1}{3x + 3}$$

E(Y | X):

$$E(Y \mid X) = -X^2 + 2X + XE(\varepsilon \mid X)$$
$$= -X^2 + 2X - \left(\frac{X}{3X + 3}\right)$$

1.d.

$$\frac{\partial}{\partial X}E(Y \mid X) = \frac{\partial}{\partial X}\left(-X^2 + 2X - \left(\frac{X}{3X+3}\right)\right)$$

$$= -2X + 2 - \frac{1}{3(X+1)^2}$$

Compared to ACE(X,ε):

$$ACE(X,\varepsilon) = \frac{5 - 6x^2}{3x+3}$$
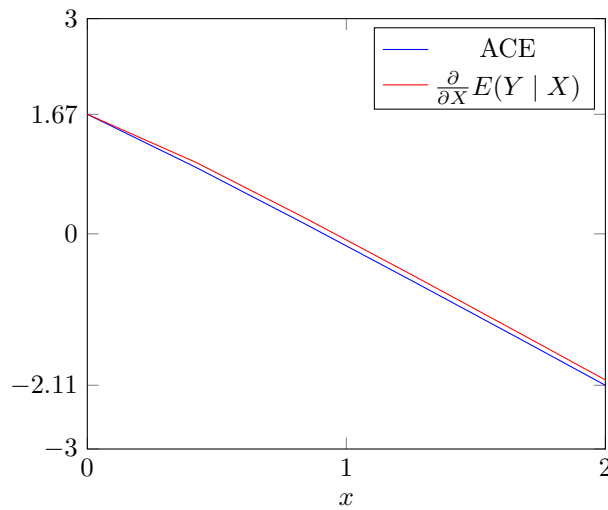
As we can see from the plot there is only a small difference between $\frac{\partial}{\partial X}E(Y \mid X)$ and ACE(X,ε).
Here it appears that $\frac{\partial}{\partial X}E(Y \mid X)$ and ACE(X,ε) are very similar over the set of possible X values.
This would suggest that our regression estimates may be interpreted as estimates of the causal impact
of X on Y.

1.e.

The BLP of Y given X solves:

$$\min_{\alpha,\beta} E(Y - \alpha - \beta X)^2$$

Solution:

$$\beta^* = \frac{Cov(X,Y)}{Var(X)}$$

$$\alpha^* = E(Y) - \beta^* E(X)$$

$E(X)$, $E(X^2)$, $E(Y \mid X)$ and $E(Y)$:

$$E(X) = \frac{7}{6}$$

$$E(X^2) = \frac{5}{3}$$

$$E(Y \mid X) = -X^2 + 2X - \left(\frac{X}{3X + 3} \right)$$

$$E(Y) = \frac{1}{2}$$

Cov(X, Y):

$$
\begin{aligned}
Cov(X, Y) &= E(XY) - E(X)E(Y) \\
&= E[X E(Y \mid X)] - E(X)E(Y) \\
&= E\left[-X^3 + 2X^2 - \left(\frac{X^2}{3X + 3} \right) \right] - \frac{7}{12} \\
&= \int_0^2 \left(-X^3 + 2X^2 - \frac{X^2}{3X + 3} \right) \cdot \left(\frac{1 + x}{4} \right) dx - \frac{7}{12} \\
&= \frac{23}{45} - \frac{7}{12} \\
&= -\frac{13}{180}
\end{aligned}
$$

Var(X, Y):

$$
\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
&= \frac{5}{3} - \frac{49}{36} \\
&= \frac{11}{36}
\end{aligned}
$$

So the BLP estimates are:

$$\beta^* = \frac{Cov(X, Y)}{Var(X)} = \frac{-\frac{13}{180}}{\frac{11}{36}} = -\frac{13}{55}$$

$$\alpha^* = E(Y) - \beta^* E(X) = \frac{1}{2} + \frac{13}{55} \cdot \frac{7}{6} = \frac{128}{165}$$

OLS of Y on constant and X estimates the BLP as the OLS estimates are the sample analogs of the first order conditions that characterise the solution to the BLP.

Question 2

2.a

$$\tilde{\beta} = \frac{\bar{Y}}{\bar{X}}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i}{\frac{1}{n}\sum_{i=1}^{n} X_i}$$

$$= \left(\frac{1}{\sum_{i=1}^{n} X_i}\right) \cdot \sum_{i=1}^{n} Y_i$$

$$= g\left(\sum_{i=1}^{n} Y_i\right)$$

This is a linear function of Y, therefore $\tilde{\beta}$ is a linear estimator.

$$E(\tilde{\beta} \mid X) = E\left(\frac{\bar{Y}}{\bar{X}} \mid X\right)$$

$$= \frac{1}{n\bar{X}} E\left(n\bar{Y} \mid X\right)$$

$$= \frac{1}{n\bar{X}} E\left(\sum_{i=1}^{n} X_i\beta + \varepsilon \mid X\right)$$

$$= \frac{1}{n\bar{X}} \left(E(\sum_{i=1}^{n} X_i\beta + \varepsilon) \mid X\right)$$

$$= \frac{1}{n\bar{X}} \left(\sum_{i=1}^{n} X_i\beta + 0\right)$$

$$= \frac{1}{n\bar{X}} \left(\beta\sum_{i=1}^{n} X_i\right)$$

$$= \beta\frac{n\bar{X}}{n\bar{X}}$$

$$= \beta$$

$\tilde{\beta}$ is conditionally unbiased.

$$Var(\tilde{\beta} \mid X) = Var\left(\frac{\bar{Y}}{\bar{X}} \mid X\right)$$

$$= \frac{1}{n^2 \bar{X}^2} Var\left(n\bar{Y} \mid X\right)$$

$$= \frac{1}{n^2 \bar{X}^2} Var\left(\sum_{i=1}^{n} Y_i \mid X\right)$$

$$= \frac{1}{n^2 \bar{X}^2} Var\left(\sum_{i=1}^{n} X_i\beta + \varepsilon \mid X\right)$$

$$= \frac{1}{n^2 \bar{X}^2} \left(\sum_{i=1}^{n} Var(X_i\beta + \varepsilon \mid X)\right)$$

$$= \frac{1}{n^2 \bar{X}^2} \left(\beta^2 \sum_{i=1}^{n} Var(X_i \mid X) + \sum_{i=1}^{n} Var(\varepsilon \mid X)\right)$$

$$= \frac{1}{n^2 \bar{X}^2} \left(0 + \sum_{i=1}^{n} \sigma^2\right)$$

$$= \frac{\sigma^2}{\bar{X} \sum_{i=1}^{n} X_i}$$

Compare to the conditional variance of the OLS estimator:

$$Var(\hat{\beta}^{OLS} \mid X) = \sigma^2 (X'X)^{-1}$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}$$

$Var(\hat{\beta}^{OLS} \mid X)$ is smaller than $Var(\tilde{\beta} \mid X)$ as $\sum_{i=1}^{n} X_i^2$ is greater than $\bar{X} \sum_{i=1}^{n} X_i$.

2.b

$$\hat{\beta}_{(m)} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{m}(x_i - \bar{x})^2}$$

$$= (X_m'X_m)^{-1}(X_m'Y_m)$$

This estimator is linear in y.

$$E[\hat{\beta}_{(m)} \mid X] = E[(X_m'X_m)^{-1}(X_m'Y_m) \mid X]$$

$$= (X_m'X_m)^{-1}X_m'E[Y_m \mid X]$$

$$= (X_m'X_m)^{-1}X_m'X_m\beta$$

$$= \beta$$

Thus, $\hat{\beta}_{(m)}$ is conditionally unbiased.

$$
\begin{aligned}
Var(\hat{\beta}_{(m)} \mid X) &= Var((X_m'X_m)^{-1}(X_m'Y_m) \mid X) \\
&= (X_m'X_m)^{-1}X_m'Var(Y_m \mid X)X_m(X_m'X_m)^{-1} \\
&= (X_m'X_m)^{-1}X_m'\sigma^2 I_m X_m(X_m'X_m)^{-1} \\
&= \sigma^2(X_m'X_m)^{-1}X_m'X_m(X_m'X_m)^{-1} \\
&= \sigma^2(X_m'X_m)^{-1}
\end{aligned}
$$

The unconditional variance of the OLS estimator using all n observations is:

$$
Var(\hat{\beta} \mid X) = \sigma^2(X'X)^{-1}
$$

Comparing variances:

$$
(X_m'X_m)^{-1} = \frac{1}{\sum_{i=1}^{m} x_i^2}
$$

$$
(X'X)^{-1} = \frac{1}{\sum_{i=1}^{n} x_i^2}
$$

$Var(\hat{\beta}_{(m)} \mid X) \geq Var(\hat{\beta} \mid X)$ as $\sum_{i=1}^{m} x_i^2 \leq \sum_{i=1}^{n} x_i^2$

2.c

A minimum conditional variance, linear estimator could simply be a constant function such as $\hat{\beta} = 1$. This is both linear and has zero variance.

Question 3

3.a

Only one observation of (y,x) available, so the OLS estimate simplifies to:

$$\hat{\beta}^{OLS} = (X'X)^{-1}X'Y$$
$$= \frac{xy}{x}$$
$$= \frac{y}{x}$$

Calculating the unconditional mean of the estimator:

$$E(\hat{\beta}^{OLS}) = E\left(\frac{y}{x}\right)$$
$$= E\left(\frac{\beta x + \varepsilon}{x}\right)$$
$$= E(\beta) + E\left(\frac{\varepsilon}{x}\right)$$
$$= \beta + E(\varepsilon) \cdot E\left(\frac{1}{x}\right)$$
$$= \beta$$

Calculating the unconditional variance of the estimator:

$$Var(\hat{\beta}^{OLS}) = Var\left(\frac{y}{x}\right)$$
$$= Var\left(\frac{\beta x + \varepsilon}{x}\right)$$
$$= Var\left(\beta + \frac{\varepsilon}{x}\right)$$
$$= E\left[\left(\beta + \frac{\varepsilon}{x}\right)^2\right] - E\left[\beta + \frac{\varepsilon}{x}\right]^2$$
$$= E\left[\beta^2 + 2\beta\frac{\varepsilon}{x} + \frac{\varepsilon^2}{x^2}\right] - \beta^2$$
$$= \beta^2 + 2\beta E\left[\frac{1}{x}\right]E[\varepsilon] + E\left[\frac{1}{x^2}\right]E[\varepsilon^2] - \beta^2$$
$$= \beta^2 + \frac{625}{49} - \beta^2$$
$$= \frac{625}{49}$$

10

3.b

$$E(\tilde{\beta}) = E(xy)$$
$$= E(\beta x^2 + \varepsilon x)$$
$$= \beta E(x^2) + E(\varepsilon x)$$
$$= \beta E(x^2) + E(\varepsilon)E(x)$$
$$= \beta(1) + 0$$
$$= \beta$$

$\tilde{\beta}$ is unconditionally unbiased.

$$Var(\tilde{\beta}) = Var(xy)$$
$$= Var(\beta x^2 + \varepsilon x)$$
$$= E[(\beta x^2 + \varepsilon x)^2] - E[\beta x^2 + \varepsilon x]^2$$
$$= E[\beta^2 x^4 + 2\beta\varepsilon x^3 + \varepsilon^2 x^2] - E[\beta x^2 + \varepsilon x]^2$$
$$= \beta^2 E[x^4] + 0 + E[\varepsilon^2]E[x^2] - \beta^2$$
$$= \beta^2 \left(\frac{1201}{625}\right) + 1 - \beta^2$$
$$= \beta^2 \left(\frac{576}{625}\right) + 1$$

3.c

When $\beta = 0$:

$$Var(\tilde{\beta}) - Var(\hat{\beta}^{OLS}) = 1 - \frac{625}{49}$$

As a result:

$$Var(\hat{\beta}^{OLS}) \geq Var(\tilde{\beta})$$

The Gauss-Markov theorem states that OLS is BLUE - i.e. has the minimum variance of all linear conditionally unbiased estimators. In this setting $\tilde{\beta}$ is an unconditionally unbiased estimator meaning that the fact that $Var(\hat{\beta}^{OLS}) \geq Var(\tilde{\beta})$ does not contradict the Gauss-Markov theorem.

Question 4

The faulty regression consisted of running:

$$x_i = \alpha + \beta y_i + \varepsilon_i$$

OLS coefficient estimates from this regression:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$\hat{SE}(\hat{\beta}) = \sqrt{\frac{\frac{1}{n-2}SSR}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{3}$$

$$\hat{SE}(\hat{\alpha}) = \sqrt{\frac{SSR \cdot \frac{1}{n}\sum_{i=1}^{n} y_i^2}{n - 2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2} \tag{5}$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-2} \cdot (1 - R^2) \tag{6}$$

$$SSR = \sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{7}$$

OLS coefficient estimates we want to obtain:

$$\hat{\beta}^* = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{8}$$

$$\hat{SE}^*(\hat{\beta}^*) = \sqrt{\frac{\frac{1}{n-2}SSR}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{9}$$

$$\hat{\alpha}^* = \bar{y} - \hat{\beta}^*\bar{x} \tag{10}$$

$$\hat{SE}^*(\hat{\alpha}^*) = \sqrt{\frac{SSR \cdot \frac{1}{n}\sum_{i=1}^{n} x_i^2}{n - 2 \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{11}$$

$$R^{2*} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{12}$$

$$\bar{R}^{2*} = 1 - \frac{n-1}{n-2} \cdot (1 - R^{2*}) \tag{13}$$

$$SSR^* = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{14}$$

From (5) and (7):

$$SSR = (1 - R^2) \sum_{i=1}^{n} (x_i - \bar{x}_i)^2 \tag{15}$$

$$\sum_{i=1}^{n} (x_i - \bar{x}_i)^2 = \frac{SSR}{1 - R^2} \tag{16}$$

Now we also have:

$$R^2 = R^{2*} \tag{17}$$

$$\bar{R}^2 = \bar{R}^{2*} \tag{18}$$

Using equation (2):

$$\hat{SE}(\hat{\beta}) = \sqrt{\frac{\frac{1}{n-2} SSR}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{19}$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{SSE}{\hat{SE}(\hat{\beta})^2 \cdot (n - 2)} \tag{20}$$

Obtaining the remaining correct estimates using the equations above:

- To get n, we can rearrange equation (6)

- From knowing n, we can calculate $\sum_{i=1}^{n} (y_i - \bar{y})^2$

- Combining $\sum_{i=1}^{n} (y_i - \bar{y})^2$ and $\sum_{i=1}^{n} (x_i - \bar{x})^2$ we can calculate $\hat{\beta}^*$ and $\hat{SE}(\hat{\beta}^*)$.

- SSR^* can now also be calculated:

$$SSR^* = (1 - R^2) \sum_{i=1}^{n} (y_i - \bar{y}_i)^2 \tag{21}$$

All that remains is $\hat{\alpha}^*$ and $\hat{SE}(\hat{\alpha}^*)$ which require $\frac{1}{n} \sum_{i=1}^{n} y_i^2$, $\bar{x}$ and $\frac{1}{n} \sum_{i=1}^{n} x_i^2$.

- $\frac{1}{n} \sum_{i=1}^{n} x_i^2$ can now be calculated via equation (4).

- $\bar{x}$ can now be calculated from $\hat{\beta}^*$ and equation (3).

- $\frac{1}{n} \sum_{i=1}^{n} x_i^2$ can now be calculated from n, $\sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\bar{x}$.

Question 5

Stata code can be found at the following link.

5.a

Results from regressing log(wage) on constant, education, experience and experience squared:

Table 1: log(wage) regressed on constant, education, experience and experience squared

	Log earnings
Education (years)	0.144***
	(0.0131)
Experience (years)	0.0265
	(0.0284)
Experience squared	0.000969
	(0.00134)
Constant	7.385***
	(0.280)
Observations	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.b

Results from regressing log(wage) on a constant, education, experience and the fitted values from part a:

Table 2: log(wage) regressed on constant, education, experience and fitted values

	Log earnings
Education (years)	-0.00000423
	(0.200)
Experience (years)	-0.00000137
	(0.0642)
Fitted values	1.000
	(1.382)
Constant	-0.000210
	(10.08)
Observations	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The obtained coefficient estimates are all approximately zero apart from the coefficient on the fitted values which is approximately 1.

We know that the fitted values arise from the regression of log(wages) on constant, education, experience and experience squared. Therefore all the variation in log(wages) can be explained by the fitted values, resulting in coefficient estimates of 0 for the constant, education and experience and 1 for the fitted values.

5.c

Results from regressing partialled out log(wage) on partialled out experience.

Table 3: Partialled out log(wage) regressed on partialled out experience

	Partialled out log(wage)
Partialled out experience	0.0265
	(0.0284)
Constant	1.12e-10
	(0.0210)
Observations	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

When partialled out, the regression coefficient estimate for the effect of experience on log(wages) has stayed the same as in the first regression (0.0265). This is implied by the Frisch-Waugh-Lovell theorem which states that the coefficient estimate from the partialled out regression will be equal to the coefficient estimate from the standard OLS regression in part a.

The standard error of this coefficient estimate has increased, resulting from a change to the degrees of freedom in the estimated regression.

The constant term estimate being approximately zero occurs due to the fact that partialled out log wages are independent of the constant.

5.d

Results from regressing log(wage) on partialled out experience are displayed below:

Table 4: log(wage) regressed on partialled out experience

	Log earnings
Partialled out experience	0.0265
	(0.0297)
Constant	9.645***
	(0.0219)
Observations	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Here the regression estimate for the coefficient on partialled out experience is the same as in the previous regression. The regression estimate for the constant has changed as this is no longer independent of log(wages).

The standard errors of the estimates have changed as the residuals have changed, due to the fact we have not partialled out log(wages) in this regression.