Quantitative Economics

1026376

A12747W1

Part A

1. A company produces light bulbs. Let Y be a random variable that takes on the value 1 if a light bulb breaks within a year of being produced, and 0 otherwise. Let the probability that the light bulb breaks within a year be p. Let $Y_1, \ldots, Y_n$ be i.i.d. draws from this distribution.

   (a) Derive the mean and variance of the sample average $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, as a function of p and n alone. [40%]

This is a Bernoulli variable with $\mathbb{E}Y = p$ and $Var(Y) = p(1-p)$

(a) $\mathbb{E}\bar{Y} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} Y_i$. By linearity of exp,

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

$$= \frac{1}{n}\left[\mathbb{E}(Y_1) + \mathbb{E}Y_2 + \cdots + \mathbb{E}Y_n\right]$$

By iid,

$$= \frac{1}{n} n \mathbb{E}Y$$

$$= p.$$

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

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

$$= \frac{1}{n^2} Var(Y_1 + Y_2 + \cdots Y_n)$$

$$= \frac{1}{n^2} Var(Y_1) + Var(Y_2) + \cdots Var(Y_n)$$

$$+ \cancel{Cov(Y_1, Y_2)} + \cancel{Cov(Y_1, Y_3)} + \cdots$$

By the independence assumption,

$$= \frac{1}{n^2} Var(Y_1) + \cdots + Var(Y_n)$$

$$\overset{iid}{=} \frac{1}{n^2} n\, Var(Y)$$

$$= \frac{p(1-p)}{n} \quad \square$$

Variance of Bernoulli RV $= p(1-p)$.

Proof:

$Var(X) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2$

$= p - p^2$ ▫

(b) Explain if anything else can be said about the distribution of [10%]

$$\frac{\bar{Y} - \mathbb{E}(\bar{Y})}{[\text{var}(\bar{Y})]^{1/2}}$$

when n is large.

Yes, by the CLT. The CLT states that if the X_is are iid, and $0 < Var(X_i) < \infty$, then

$$\frac{\bar{Y} - \mathbb{E}\bar{Y}}{SE(\bar{Y})} \sim N(0,1). \quad \text{as} \quad n \longrightarrow \infty.$$

(c) Suppose you buy $n = 100$ light bulbs from this company. If $p = 0.2$, how likely is it that the number of lightbulbs that no longer work after one year is between 15 and 25 (inclusive)? [Hint: you may use the result from part (b) to simplify your calculations.] [50%]

If the number of lightbulbs breaking is between 15 and 25, that means we are looking for $Pr[0.15 \leq \bar{Y} \leq 0.25]$.

From the result in (b), we know that

$$t = \frac{\bar{Y} - \mathbb{E}(\bar{Y})}{SE(\bar{Y})} \sim N(0,1) \quad \text{since } n = 100$$

which is "large enough" to use the CLT approx.

We can further decompose this into

$$Pr[\bar{Y} \leq 0.25] - Pr[\bar{Y} \leq 0.14]$$

and calculate them one by one. Standardising and substituting, we have

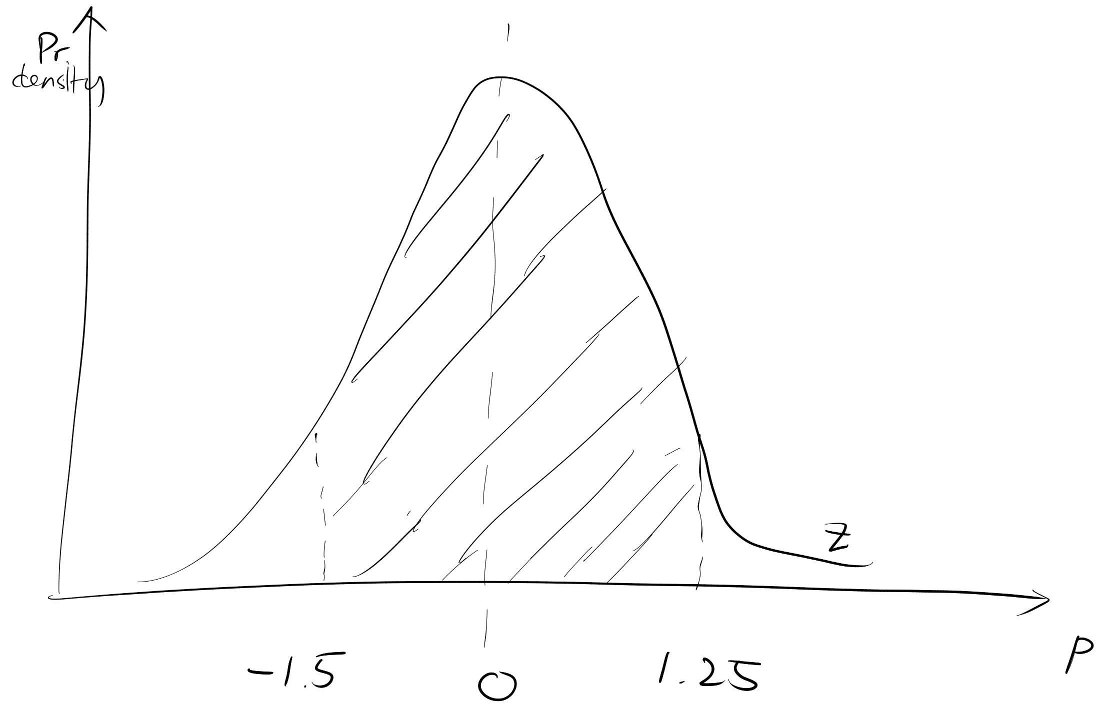
$$Pr\left[\frac{\bar{Y} - 0.20}{\sqrt{\frac{(0.2)(0.8)}{100}}} \leq \frac{0.25 - 0.20}{\sqrt{\frac{(0.2)(0.8)}{100}}}\right] - Pr\left[\frac{\bar{Y} - 0.20}{\sqrt{\frac{(0.2)(0.8)}{100}}} < \frac{0.14 - 0.20}{\sqrt{\frac{(0.2)(0.8)}{100}}}\right]$$

$$\Rightarrow Pr[N(0,1) \leq 1.25] - Pr[N(0,1) < -1.5]$$

$$= 0.894 - 0.0668$$

$$= 0.828.$$

Here's a graph of what that looks like :



2. The mean squared forecast error (MSFE) of a variable Y_{t+1}, using the predictor X_t, is defined as $\text{MSFE}(X_t) = \mathbb{E}(Y_{t+1} - X_t)^2$. Suppose that the only information available in period t is Y_t itself (i.e. $Y_{t-1}, Y_{t-2}, \ldots$ are not observed).

(a) What is the MSFE-minimising forecast of Y_{t+1}, using the information available in period t? Provide an explanation (it is not sufficient to merely state an answer). [50%]

Claim: The MSFE - minimising forecast of Y_{t+1} using only Y_t is the CEF:

$$\mathbb{E}[Y_{t+1} \mid Y_t]$$

Proof: The CEF provides the best predictor of Y_{t+1} using Y_t alone. Here is why:

Because we only have Y_t, we have that MSFE is minimised when

$$MSFE_{m(Y_t)} = \mathbb{E}[Y_{t+1} - m(Y_t)]^2$$ where $m(\cdot)$ is any function of $\cdot$ alone.

We can rewrite this as

$$\mathbb{E}\Big[\big\{\underbrace{(Y_{t+1} - \mathbb{E}[Y_{t+1} \mid Y_t])}_{(a)} - \underbrace{(m(Y_t) - \mathbb{E}[Y_{t+1} \mid Y_t])}_{(b)}\big\}^2\Big]$$

Define $\varepsilon :=$ (a). and $g(Y_t) :=$ (b), the latter being a function of Y_t alone. Then we have $\mathbb{E}\big[\{\varepsilon + g(Y_t)\}^2\big]$.

Expanding, and by linearity, we have

$$\Rightarrow \mathbb{E}\varepsilon^2 + \underbrace{2\mathbb{E}(\varepsilon g(Y_t))}_{(2)} + \mathbb{E}\{g(Y_t)\}^2.$$

Rewrite (2) as follows:

$$\mathbb{E}(\varepsilon g(Y_t)) \overset{LIE}{=} \mathbb{E}(\mathbb{E}(gY_t)\varepsilon \mid Y_t)) \overset{\text{conditioning}}{=} \mathbb{E}[g(Y_t)\, \mathbb{E}(\varepsilon \mid Y_t)]$$

But by the CEF decomposition we know we can decompose Y_{t+1} as

$$Y_{t+1} = \mathbb{E}[Y_{t+1} | Y_t] + \varepsilon \quad \text{where } \mathbb{E}[\varepsilon / Y_t] = \mathbb{E}(\varepsilon) = 0.$$

Thus, we have that $\mathbb{E}[g(Y_t)\cancel{\mathbb{E}(\varepsilon / Y_t)}^{=0}] = 0$, and thus

$$MSFE_{m(Y_t)} = \mathbb{E}\varepsilon^2 + \mathbb{E}\{g(Y_t)\}^2.$$

This MSFE is minimised precisely when $g(Y_t) = 0$,
That is, when

(b)

$$m(Y_t) = \mathbb{E}[Y_{t+1} / Y_t].$$

Hence, the MSFE minimising forecast using only Y_t alone is the conditional expectation function $\mathbb{E}[Y_{t+1} | Y_t]$ □.

(b) Under what conditions is the MSFE-minimising forecast in part (a) equal to a constant plus Y_t? Give an example of such a variable. [50%]

In part a) we proved that the MSFE minimising forecast is the CEF $\mathbb{E}[Y_{t+1} | Y_t]$.

The question is essentially asking when the CEF is linear, AND in the following form:

$$\mathbb{E}[Y_{t+1} | Y_t] = \alpha + Y_t.$$

This means that the process must have a unit root and have a deterministic trend.

An example would be a random walk with a deterministic trend.

Suppose

$$Y_{t+1} = Y_t + \alpha + u_{t+1}. \quad \text{where} \quad u_{t+1} \perp Y_t, \alpha,$$

Then the CEF would be and $\mathbb{E}u_t = 0 \ \forall t$.

$$\mathbb{E}[Y_{t+1} | Y_t] = \mathbb{E}[Y_t + \alpha + u_{t+1} | Y_t]$$

$$= \mathbb{E}Y_t|Y_t + \mathbb{E}\alpha|Y_t + \mathbb{E}[u_{t+1}|Y_t]$$

$$\overset{\text{conditioning,}}{=} Y_t + \alpha + \mathbb{E}[u_{t+1}|Y_t]$$

$$\overset{\text{independence,}}{=} Y_t + \alpha + \cancel{\mathbb{E}u_{t+1}}^{=0}$$

$\square$.

3. You have data on wages (W_i; in £/hr), years of experience (X_i), and the area in which an individual lives, which is classified as either a city, a town, or a rural area. Define C_i to be the dummy variable that takes the value 1 if individual i lives in a city, and 0 otherwise; analogously define T_i and R_i for towns and rural areas respectively. Consider the following two regression models

$$W_i = \beta_0 + \beta_X X_i + \beta_C C_i + \beta_T T_i + u_i \tag{1}$$
$$W_i = \gamma_0 + \gamma_X X_i + \gamma_C C_i + \gamma_R R_i + v_i \tag{2}$$

where $\mathbb{E}[u_i \mid X_i, C_i, T_i] = 0$ and $\mathbb{E}[v_i \mid X_i, C_i, R_i] = 0$.

(a) Suppose you estimate both (1) and (2) by ordinary least squares (OLS). [50%]

(i) Show that the sample covariance between the OLS residuals $\hat{u}_i$ (obtained from model (1)) and R_i is zero.

(ii) Derive the relationship between the estimates $(\hat{\beta}_X, \hat{\beta}_C, \hat{\beta}_T)$ and $(\hat{\gamma}_X, \hat{\gamma}_C, \hat{\gamma}_R)$.

[Hint: use the fact that an OLS regression uniquely decomposes the dependent variable into a linear function of the regressors, and a residual that has zero sample mean and zero sample covariance with each of the regressors.]

(i) Claim: $\widehat{Cov}(\hat{u}_i, R_i) = 0$.

Proof: Because one individual can live only in one place (otherwise we wouldn't have to avoid the dummy variable trap), we know that R_i is a deterministic function of T_i and C_i.

Specifically,

$$R_i = (1-T_i)(1-C_i)$$

So we can rewrite the following:

$$\widehat{Cov}(\hat{u}_i, R_i) = \widehat{Cov}(\hat{u}_i, (1-T_i)(1-C_i))$$

Expanding, $= \widehat{Cov}(\hat{u}_i, 1 - T_i - C_i + C_iT_i)$

BUT because T_i and C_i can never both equal 1,

$$= \widehat{Cov}(\hat{u}_i, 1 - T_i - C_i)$$

Further, by construction in the pop LR in (1)

$$Cov(u_i, T_i) = 0 \text{ and } Cov(u_i, C_i) = 0.$$

By linearity of covariance we thus get 0, and hence

$$\widehat{Cov}(\hat{u}_i, R_i) = 0 \quad \square.$$

(ii) Derive the relationship between the estimates $(\hat\beta_X, \hat\beta_C, \hat\beta_T)$ and $(\hat\gamma_X, \hat\gamma_C, \hat\gamma_R)$.

[Hint: use the fact that an OLS regression uniquely decomposes the dependent variable into a linear function of the regressors, and a residual that has zero sample mean and zero sample covariance with each of the regressors.]

(ii) We may write W as (dropping subscripts for ease of writing),

$$W = \hat\beta_0 + \hat\beta_X X + \hat\beta_C C + \hat\beta_T T + e_1 \quad (\text{pop } LR(1))$$

and

$$W = \hat\gamma_0 + \hat\gamma_X X + \hat\gamma_C C + \hat\gamma_R R + e_2 \quad (\text{pop } LR\,2).$$

where e_1 and e_2 are mean zero and OR.

$$\hat\beta_X = \frac{\hat{Cov}(W, \tilde X_1)}{\hat{Var}(\tilde X_1)}, \quad \hat\gamma_X = \frac{\hat{Cov}(W, \tilde X_2)}{Var(\tilde X_2)}$$

where $X = \pi_0 + \pi_1 C + \pi_2 T + \tilde X_1$ (a) and $X = \delta_0 + \delta_1 C + \delta_2 R + \tilde X_2$ (b)

But recall again that $R = (1-T)(1-C)$, and we can rewrite (b) as

$$X = \delta_0 + \delta_1 C + \delta_2 (1-T)(1-C) + \tilde X_2$$
$$= \delta_0 + \delta_1 C + \delta_2 (1 - C - T + TC) + \tilde X_2$$
$$= \delta_0 + \delta_1 C + \delta_2 - \delta_2 C - \delta_2 T + \delta_2 TC + \tilde X_2$$
$$X = (\delta_0 + \delta_2) + (\delta_1 - \delta_2) C - \delta_2 T + \delta_2 TC + \tilde X_2$$

And because T and C can't ever be 1 at the same time, we can just strike the term off:

$$X = (\delta_0 + \delta_2) + (\delta_1 - \delta_2) C - \delta_2 T + \tilde X_2.$$

Compare with (a):

$$X = \pi_0 + \pi_1 C + \pi_2 T + \tilde X_1$$

Because a pop LR gives us a unique solution, this implies that

$\tilde{X}_1 = \tilde{X}_2$, and therefore

$$\Rightarrow \quad \hat{\beta}_X = \hat{\gamma}_X \qquad \square$$

We can repeat a similar procedure mutatis mutandis for c again.

$$\hat{\beta}_C = \frac{\widehat{Cov}(W, \tilde{C}_1)}{\widehat{Var}(\tilde{C}_1)}, \quad \hat{\gamma}_C = \frac{\widehat{Cov}(W, \tilde{C}_2)}{Var(\tilde{X}_2)}$$

and this will also give us

$$\hat{\beta}_C = \hat{\gamma}_C.$$

Finally let's compare $\hat{\beta}_T$ and $\hat{\gamma}_R$:

$$\hat{\beta}_T = \frac{\widehat{Cov}(W, \tilde{T})}{\widehat{Var}(\tilde{T})}, \quad \hat{\gamma}_R = \frac{\widehat{Cov}(W, \tilde{R})}{Var(\tilde{R})}$$

We have $T = \pi_0 + \pi_1 X + \pi_2 C + \tilde{T}$ (a) and

$R = \delta_0 + \delta_1 X + \delta_2 C + \tilde{R}$ (b)

Again write R as $(1-T)(1-C)$ and simplify:

from (b)

$$1 - T - C = \delta_0 + \delta_1 X + \delta_2 C + \tilde{R}$$

$$-T = \delta_0 - 1 + \delta_1 X + (\delta_2 + 1) C + \tilde{R}$$

$$T = \eta_0 + \eta_1 X + \eta_2 C - \tilde{R}_2$$

Compare again with (b)

$$T = \delta_0 + \delta_1 X + \delta_2 C + \tilde{R}.$$

Again because the pop LR uniquely decomposes T, all coeffs are the same and thus

note minus sign

$$\tilde{R} = -\tilde{R}_2.$$

Hence, $\hat{\beta}_T = -\hat{\gamma}_R$. $\square$

(b) Explain how you could use this data to test the hypothesis that the return to experience (the effect of X on W) is the same in cities, towns and rural areas. [50%]

You'd do an F-test. As I've shown in a ii),

the coefficients are identical so which one you use

doesn't matter. Take (1):

$$W = \beta_0 + \beta_X X + \beta_C C + \beta_T T + e_1 \quad (\text{pop } LR(1))$$

$$H_0: \beta_C = \beta_T = 0.$$

$$H_1: \beta_C \neq 0 \text{ or } \beta_T \neq 0.$$

Why this test? If the returns to living are the same everywhere, then they shouldn't affect wages and so the coeffs should equal zero.

We would thus estimate the restricted and unrestricted models

$$(R): W_i = \pi_0 + \pi_1 X_i + u_{R_i},$$

$$(UN): W_i = \beta_0 + \beta_X X_i + \beta_C C + \beta_T T + u_{UN_i}$$

and compare $SSR_{rs} - SSR_{un}$ where $SSR = \sum_{i=1}^{n} \hat{u}_i^2$

Part B

4. Suppose you work in the Human Resources department of a large company, which employs 1,000 call centre employees. About two years ago, the company offered an on-the-job training programme to all 1,000 call centre employees. 100 employees chose to participate in the training, while the remaining 900 chose not to.

   (a) Your records indicate that the employees who participated in the training now earn £3,800 per month (on average) with a sample standard deviation of 750, while the employees who did not participate earn £3,200 per month with a sample standard deviation of 750. Test whether the mean earnings of the trained employees are different from the mean earnings of the untrained employees at the 5% level. Clearly state any assumptions and results that your test relies on. [20%]

Let the mean earnings of the employees with training be μ_T and those without be μ_N.

Let the sample mean earnings be $\bar{Y}_T$ and $\bar{Y}_N$ respectively, and Y_i the earnings of an employee i.

1. Set up the null and alternative hypotheses:

$$H_0: \mu_N = \mu_T.$$

$$H_1: \mu_T \neq \mu_N.$$

2. Under the null, the distribution of the test statistic, assuming that Y_i are iid and $0 < var(Y) < \infty$, is by the CLT

$$t = \frac{\bar{Y}_T - \bar{Y}_N}{SE(\bar{Y}_T - \bar{Y}_N)} \sim N(0,1).$$

3. At the 5% level, the critical values are ± 1.96, and the decision rule is

DR: Reject H_0 if $|t^{act}| > 1.96$.

4. Find t^{act},

$$t = \frac{\bar{Y}_T - \bar{Y}_N}{se(\bar{Y}_T - \bar{Y}_N)} = \frac{\bar{Y}_T - \bar{Y}_N}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_N^2}{n_N}}}$$

$$t^{act} = \frac{3800 - 3200}{\sqrt{\frac{750^2}{100} + \frac{750^2}{900}}}$$

we are given $sd(\bar{Y}_{T/N}) = \sqrt{s^2_{T/N}}$

$$= 7.59$$

5. Since $|t^{act}| = 7.59 > 1.96$, by our DR we reject the null that the average wages are equal at the 5% Level.

(b) After performing the calculations in part (a), you start to wonder whether the difference in average earnings between the two groups can be interpreted as the causal effect of the training programme on earnings. Explain how the difference in average earnings can be decomposed into a causal effect and a selection bias term. Why might selection bias be a problem in the present setting? In what circumstances is the selection bias likely to be small? [15%]

Explain how the difference
in average earnings can be decomposed into a causal effect and a selection bias
term.

The difference in earnings can be written as

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$$

where D_i is whether or not the training was taken up. If we allow for a causal model

$$Y_i = \beta_{0i} + \beta_{1i} D_i,$$ then substituting

gives us

$$\mathbb{E}[\beta_{0i} + \beta_{1i} D_i \mid D_i = 1] + \mathbb{E}[\beta_{0i} + \beta_{1i} \overset{=0}{D_i} \mid D_i = 0]$$

by linearity
$$= \mathbb{E}[\beta_{0i} \mid D_i = 1] - \mathbb{E}[\beta_{0i} \mid D_i = 0] + D_i[\mathbb{E}\, \beta_{1i} \mid D_i = 1]$$

Since D_i is a binary var

$$= \underbrace{\mathbb{E}[\beta_{1i} \mid D_i = 1]}_{\text{TOT.}} + \underbrace{\mathbb{E}[\beta_{0i} \mid D_i = 1] - \mathbb{E}[\beta_{0i} \mid D_i = 0]}_{\text{SB}}$$

In other words, the effect of training on earnings can be decomposed into the sum of the average causal effect of the training program on those who were treated, (TOT), and the selection bias: the difference in untreated outcomes between those who took up the training program and those who didn't.

Why might selection bias be a problem in the present setting? In what circumstances is the selection bias likely to be small?

In the present setting, participation is voluntary. Suppose that hardworking people are more likely to choose to take part in the training program, and they were also more likely to earn more money (b/c they work harder). Then, there would be a diff.

in their wages even if they hadn't participated

in the training program ; $\mathbb{E}[\beta_{0i} \mid D=1] > \mathbb{E}[\beta_{0i} \mid D=0]$.

Hence selection bias is a problem here.

The selection bias is likely to be small if factors that affect your earnings are not very correlated with the likelihood of you attending the training program : say if the programme were only an hour long and very easy to do.

(c) Explain how a randomised controlled trial (RCT) can help you overcome the selection problem. [10%]

Since assignment to the treatment is randomised,

D is independent of all possible factors that could conceivably affect earnings. That is,

$D \perp \beta_{0i}, \beta_{1i}$, and hence,

$$\mathbb{E}[\beta_{0i} \mid D_i = 1] = \mathbb{E}[\beta_{0i} \mid D_i = 0] = \mathbb{E}\beta_{0i}$$

and the selection bias would be zero.

(d) To conduct such a trial, you choose $N = 200$ call centre employees who recently joined the company. You randomly allocate $n_T = 50$ employees to a treatment group and $n_C = 150$ employees to a control group. All employees in the treatment group receive the training, while all employees in the control group do not. Two years later, treated employees earn £3,625 per month with a sample standard deviation of 750, while employees in the control group earn £3,400 with a sample standard deviation of £750. Can you now conclude that, as a result of the training, the employees have significantly different earnings at the 5% level? [5%]

I won't go through the entire hypothesis testing procedure again. Instead, here is t^{act}:

$$t = \frac{\bar{Y}_T - \bar{Y}_N}{se(\bar{Y}_T - \bar{Y}_N)} = \frac{\bar{Y}_T - \bar{Y}_N}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_N^2}{n_N}}}$$

$$t^{act} = \frac{3625 - 3400}{\sqrt{\frac{750^2}{50} + \frac{750^2}{150}}}$$

$$= \frac{225}{122.5}$$

$$= 1.84$$

$\Rightarrow$ Since $t^{act} < 1.96$, we don't reject the null that employees have significantly different average earnings after the training at the 5% s.f level.

(e) Can you use the results from part (d) to estimate the causal effect that appears in the decomposition referred to in part (b)? Explain. [10%]

As mentioned, the average treatment effect

$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$ can be written as.

$$= \underbrace{\mathbb{E}[\beta 1_i \mid D_i = 1]}_{TOT} + \underbrace{\mathbb{E}[\beta_{0i} \mid D_i = 1] - \mathbb{E}[\beta_{0i} \mid D_i = 0]}_{SB}$$

And because D is independent, $SB = 0$ and we are left with the TOT.

$\mathbb{E}[\beta 1_i \mid D_i = 1]$. But here we have perfect compliance. If you had dropouts (never-takers) would wonder if $\beta 1$ was somehow correlated with D (e.g. lazy workers don't complete the training).

Thus, $\mathbb{E}[\beta 1_i \mid D_i = 1] = \mathbb{E}\,\beta 1_i$

which recovers the average treatment effect on new workers.

External validity is a problem here.

In other words, while we have the ATE on the 200 new workers, the effect of the training might be different for ~~old~~ existing workers.

(f) Imagine that you had instead allocated $n_T = 80$ employees to the treatment group and $n_C = 120$ employees to the control group, and that again you had found that treated employees earn £3,625 per month with a sample standard deviation of 750, while employees in the control group earn £3,400 with a sample standard deviation of £750. Could you have concluded that, as a result of the training, the employees have significantly different earnings at the 5% level? Comparing your answer to your answer in part (d), can you draw any general conclusions from this example? [10%]

$$t = \frac{\bar{Y}_T - \bar{Y}_N}{se(\bar{Y}_T - \bar{Y}_N)} = \frac{\bar{Y}_T - \bar{Y}_N}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_N^2}{n_N}}}$$

$$t^{act} = \frac{3625 - 3400}{\sqrt{\frac{750^2}{80} + \frac{750^2}{120}}}$$

$$= \frac{225}{108.3}$$

$$= 2.078$$

$\Rightarrow$ Reject H_0 at 5% level.

General conclusions: for maximal power, both treatment and nontreatment groups should be the same size.

(g) Formally show that the value of the actual t statistic is greatest when the two groups in the trial (as described in parts (d) and (f)) are of equal size, given that the average difference in earnings between the two groups is £225, and the sample standard deviation of earnings is £750 in each group. [15%]

Well, all we need to show here is that

$$t^{act} = \frac{225}{\sqrt{\frac{750^2}{A-x} + \frac{750^2}{x}}} \quad \text{is maximised when } x = \frac{A}{2}$$

$$0 \leq x \leq A$$

And that means finding x such that

$$\sqrt{\frac{750^2}{A-x} + \frac{750^2}{x}}$$ is minimised. Call this U,

WLOG we can square U and thus

$$\underset{x}{\operatorname{argmax}} \left(\frac{750^2}{A-x} + \frac{750^2}{x} \right) \quad \text{s.t.} \quad 0 < x \leq A.$$

By differentiating we obtain

$$\frac{\partial U}{\partial x} = -\frac{(750)^2}{(A-x)^2} - \frac{750^2}{x^2} = 0$$

$$\Rightarrow \quad (750)^2 (A-x)^2 + (750^2)(x^2) = 0$$

$$\Rightarrow \quad -(A-x)^2 = x^2$$

$$\Rightarrow \quad -(A^2 - 2AX + X^2) = x^2$$

$$\Rightarrow \quad -A^2 + 2AX = 0$$

$$\Rightarrow \quad X = \frac{A}{2}. \qquad \square.$$

You should also check SOCs but I have no time.

(h) Another company proposes to introduce a similar on-the-job training programme for its call centre employees. On the basis of the results above, how might you predict the earnings of its employees will be affected by the programme? Discuss any qualifications or caveats that you would attach to this prediction. [15%]

You'd predict that the employees' earnings would increase by £~~250~~ 225 on average. Big caveats: external validity (RCT run on new employees as mentioned — effect may be smaller for existing employees.

Also, the training programme that would be given to another company's workforce is presumably not the same as this one).

5. Suppose that Y and X are generated according to the model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u \tag{3}$$

where $\mathbb{E}[u \mid X] = 0$. Consider also the (population) linear regression model

$$Y = \gamma_0 + \gamma_1 X + v \tag{4}$$

where $\mathbb{E}v = 0$ and $\mathbb{E}Xv = 0$.

(a) Is it possible that $\mathbb{E}[v \mid X] = 0$ also? Explain. [20%]

(a) Yes. We know given the model that

$$v = \beta_2 X^2 + u. \quad \text{where} \quad \mathbb{E}[u/X] = 0.$$

Then taking conditional exps,

$$\mathbb{E}[v/X] = \mathbb{E}[\beta_2 X^2 \mid X] + \underbrace{\cancel{\mathbb{E}[u \mid X]}}_{=0}$$

$$\overset{\text{conditioning}}{=} X^2 \mathbb{E}[\beta_2 \mid X].$$

We know from the model that $\beta_2 \perp X$,

and so

$$= X^2 \mathbb{E}\beta_2$$

$$= X^2 \beta_2.$$

Hence $\mathbb{E}[v/X] = 0$ only if $\beta_2 = 0$.

(b) Two researchers, Alice and Bob, independently draw data from distinct populations, and obtain the following OLS estimates of models (3) and (4): [20%]

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_0$	$\hat{\gamma}_1$
Alice	-0.01	4.99	-1.98	-1.86	4.81
	(0.02)	(0.02)	(0.01)	(0.08)	(0.19)
Bob	0.03	5.00	-2.02	-0.08	1.00
	(0.02)	(0.03)	(0.01)	(0.21)	(0.18)

How might you account for Alice and Bob's findings? (In your answer, comment on both the point estimates and the standard errors)

The standard errors for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are quite small, and both Alice and Bob's results agree with one another. This suggests that the term u in the causal model is likewise very small.

The difference in the point estimates can be explained if Alice observed $\mathbb{E}X \approx 2.5$ and Bob observed $\mathbb{E}X \approx 1$. Why is this? Well, we know the $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ quite well, and so

$$Y = 0 + 5X - 2X^2$$ are the most likely estimates for the coeffs of the model. Then we have the short regs:

Alice: $-1.86 + 4.81X$. and

Bob: $-0.08 + X$.

By the OVB formula, $\hat{\gamma}_1$ can be given:

$$\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\pi}_1$$

where $X^2 = \pi_0 + \hat{\pi}_1 X + e$, and

it's clear that $\hat{\pi}_1 = X$.

We can see that if Bob observed $\mathbb{E}X \approx 1$,

$\mathbb{E}Y = 5\mathbb{E}X - 2\mathbb{E}X^2 = 1$ in the causal model

which would give him the point estimate

$$\hat{\gamma}_0 = \mathbb{E}Y - \hat{\gamma}_1 \mathbb{E}X.$$

$$= 0.$$

$\hat{\gamma}_1 = \beta_1 + \beta_2 \hat{\pi}_1 = 5 - 2X$

1 1

Similarly for Alice, if she observed $\mathbb{E}X \approx 0.1$

$\hat{\gamma}_1 = 4.81$ by OVB : $\hat{\gamma}_1 = \beta_1 + \beta_2 X$

5 -2

$$\hat{\gamma}_0 = \mathbb{E}Y - \hat{\gamma}_1 \mathbb{E}X$$

$$= 0.5 - 4.81(0.1)$$

$$\approx -1.86.$$

$\mathbb{E}Y = 5\mathbb{E}X - 2\mathbb{E}X^2$

$\approx 5(0.1)$

Made mistake.

(c) Compute the marginal effect of X on Y when $X = x$, in each of models (3) and (4). Show that these two effects agree when [20%]

$$x = \frac{\mathrm{cov}(X^2, X)}{2\,\mathrm{var}(X)}.$$

Marginal effects are given when $\frac{\partial Y}{\partial X}$.

In model 3, $\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 x$

4, $\frac{\partial Y}{\partial X} = \gamma_1$.

We know once again by the OVB formula that

$$\gamma_1 = \beta_1 + \beta_2 \pi_1$$

where π_1 is the coeff of X on a pop LR

$$X^2 = \pi_0 + \pi_1 X_1 + u.$$

Hence,

$\gamma_1 = \beta_1 + 2\beta_2 x$ can be rewritten as

$$\Rightarrow \cancel{\beta_1} + \beta_2 \pi_1 = \cancel{\beta_1} + 2\beta_2$$

$$\Rightarrow \pi_1 = 2x$$

$$\Rightarrow x = \frac{\mathrm{Cov}(X, X^2)}{\mathrm{Var}(X)} \left(\frac{1}{2}\right) \square.$$

(d) For the purposes of estimating the relationship between Y and X, evaluate and compare the practices of: [20%]

(i) estimating the linear model (4) and treating it as an 'approximation' to (3); and

(ii) estimating a fifth-order polynomial model, by regressing Y on X, X^2, X^3, X^4 and X^5 (and a constant).

(d) This is basically the practice of using a pop LR as a linear approx to the CEF. We can usually do this, but in this case as $\beta_2 = -2$ the approximation will be very poor when $|X|$ gets large. We could do this if $|\beta_2| \simeq 0$, or X was small.

(ii) In general you would run into problems of overfitting here if Y depended on variables other than X. But I think it would actually be OK here. You could estimate the 5^{th} order poly and do F-tests to drop the higher-order coeffs.

You would like to estimate the parameters of (3), but are concerned by the possibility that $\mathbb{E}[u \mid X] \neq 0$. Suppose you have an 'instrument' Z that satisfies

$$X = \pi_0 + \pi_1 Z + \varepsilon$$

with Z being independent of both ε and u.

(e) Show that (β_1, β_2) can be recovered from a population linear regression of Y on Z *and* Z^2 (and a constant), and of X on Z (and a constant), under suitable conditions on (π_0, π_1). [Hint: start by evaluating $\mathbb{E}[Y \mid Z]$ and $\mathbb{E}[X \mid Z]$.] [20%]

We have in the causal model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u.$$

Taking conditional expectations,

$$\mathbb{E}[Y \mid Z] = \mathbb{E}[\beta_0 + \beta_1 X + \beta_2 X^2 + u \mid Z]$$

$$= \mathbb{E}\beta_0 + \mathbb{E}\beta_1[X \mid Z] + \mathbb{E}\beta_2[X^2 \mid Z] + \underbrace{\mathbb{E}[\cancel{u} \mid Z]}_{=0}$$

Because Z is independent of X and u,

$$= \beta_0 + \beta_1 \mathbb{E}X + \beta_2 \mathbb{E}X^2.$$

$$\mathbb{E}[Y \mid Z] = \mathbb{E}[\beta_0 + \beta_1(\pi_0 + \pi_1 Z + e) + \beta_2(\pi_0 + \pi_1 Z + e)^2 \mid Z]$$

$$= \mathbb{E}[\eta_0 - \eta_1 Z + \eta_2 Z^2 \mid Z]$$

by conditioning

$$= \eta_0 + \eta_1 Z + \eta_2 Z^2$$

$\Rightarrow$ The CEF is linear in Z & Z^2. Hence, a pop LR of Y on Z and Z^2 will recover η_1, η_2.

We also have

$$\mathbb{E}[X|Z] = \mathbb{E}[\pi_0 + \pi_1 Z + \varepsilon | Z].$$

Assuming that π_0 and $\pi_1 \perp Z$,

$$= \pi_0 + \pi_1 Z + \mathbb{E}[\varepsilon | Z]$$ by independence

$$= \pi_0 + \pi_1 Z + \varepsilon.$$

$\Rightarrow$ Because the CEF is linear, a population linear regression of X on Z recovers π_0 and π_1.

We have the pop LRs:

$$\beta_1 = \frac{\mathrm{Cov}(X\ Y)}{\mathrm{Var}(X)} = \frac{\mathrm{Cov}(Y, Z)}{\mathrm{Var}(Z)} \cdot \frac{\mathrm{Var}(Z)}{\mathrm{Cov}(X, Z)}$$

$$= \eta_1 \cdot \frac{1}{\pi_1}$$

. Hence we can recover β_1, from the two Pop LRs.

$$\beta_2 = \frac{\mathrm{Cov}(X^2, Y)}{\mathrm{Var}(X^2)} = \frac{\mathrm{Cov}(Y, Z^2)}{\mathrm{Var}(Z^2)} \cdot \frac{\mathrm{Var}(Z)}{\mathrm{Cov}(X, Z)}$$

Ran out of time.