ECMA 31360, Pset 2: Using OLS to Estimate Treatment Effects (100p)

Derek Li, Andi Liu, Tessie Dong

Objective of the PSet: In PSet1 you considered the OLS estimator in a prediction context: you had a linear-in-parameters CEF, $E[y_i|D_i] = \alpha + \beta D_i$ with D_i being a 0/1 variable, and showed that the OLS estimator of β is unbiased for $E[y_i|D_i = 1] - E[y_i|D_i = 0]$. Here you consider the OLS estimator in a causal context: you start with the outcome equation[1] $y_i = \alpha + \beta D_i + u_i$, where D_i and u_i are determinants of y_i (observed and unobserved respectively), and study the circumstances under which the OLS estimator of β enables inference about the causal effect of D_i on y_i. Comparing the two contexts yields a lot of learning. To answer PSet2's questions you do <u>not</u> need the course material unless explicitly stated because Pset2 you use the traditional (i.e., pre Rubin Causal Model) approach to causal analysis (from introductory econometrics courses).

Background for the PSet: Walmart Inc. is an American retail corporation that operates a chain of hypermarkets. Walmart introduced *Sam's Club Plus* (link) in February 2018. Membership in *Sam's Club Plus* implies that customers earn cash rewards (e.g., they get $10 back for every $500 spent on qualifying purchases), enjoy free-shipping on many items, and reduced 2-day shipping charges. *Sam's Club Plus* charges an annual fee of $100. Shoppers may use brick-and-mortar Walmart stores, or shop online at Walmart.com. The questions below focus on the causal impact of *Sam's Club Plus* membership on online spend.[2]

---

[1]Think of this equation as stemming from some theoretical model which we have left unspecified.

[2]All references to Walmart.com and its customers in this problem set are entirely fictitious and are used solely for the purpose of illustrating statistical concepts and their application in various contexts.

Q1. (50p) Let y_i be customer i's spend at Walmart.com in a given month. Let $D_i = 1$ if customer i is a *Sam's Club Plus* member, $= 0$ otherwise. Assume membership status does not vary during the month. y_i is determined by the customer's membership status (D_i) and other determinants (u_i) according to the homogeneous treatment effects model:

$$y_i = \alpha + \rho D_i + u_i. \tag{1}$$

For example, u_i may include household income and size. As (y_i, D_i, u_i) vary across customers, we think of them as RVs. Let $E[u_i] = 0$, where the expectation is taken with respect to the distribution of u_i in the population of Walmart.com customers. You have data on a sample of size n of customers: $\{(y_i, D_i) | i = 1, \ldots, n\}$. Note that u_i is not included in the data for any of the sample customers.[3] Some of the sample customers are *Sam's Club Plus* members, some are not. As your data contains a mix of both types of customers we say that you have "observational variation in the cause or treatment". (α, ρ) are unknown parameters. Let $\bar{y}^0$ (respectively, $\bar{y}^1$) denote the sample average of y_i across sample customers with $D_i = 0$ (respectively, with $D_i = 1$).

(a) (2p) Provide additional examples of determinants of spend that may be part of u_i. **Hint**: In the background there is a consumer demand model, i.e., you think of model (1) as a consumer expenditure function from microeconomics.

> **Solution:** One determinant of spend that may be part of u_i is the customer's age. Younger customers may spend more than older customers, for example. Another determinant of spend that may be part of u_i is the customer's location - customers in urban areas may spend more than customers in rural areas due to the convenience of online shopping, shipping times/rates, etc.

(b) (2p) Show that ρ is the causal impact of *Sam's Club Plus* membership on a customer's spend, in the sense that, ρ is the difference in a customer's spend with and without membership holding all else the same.

> **Solution:** The causal effect of D_i on y_i, holding u_i constant, is given by:
>
> $$\frac{dy_i}{dD_i} = \frac{d(\alpha + \rho D_i + u_i)}{dD_i} = \rho$$
>
> Given the model above, we can analyze the two cases for D_i.
> When $D_i = 1$, the equation becomes:
>
> $$y_i = \alpha + \rho + u_i$$
>
> And when $D_i = 0$, the equation becomes:
>
> $$y_i = \alpha + u_i$$
>
> Now, let's consider the difference between the spend of a customer who is a Sam's Club Plus member ($D_i = 1$) and one who is not ($D_i = 0$). Holding all else the same, we can take the difference of the two above equations:
>
> $$\bar{y}^1 - \bar{y}^0 = (\alpha + \rho + u_i) - (\alpha + u_i) = \rho$$
>
> This demonstrates that ρ is the difference in a customer's spend with and without Sam's Club Plus membership, while holding all other factors constant - the unobserved term u_i cancels out in the difference. Thus, ρ in this model represents the causal impact of a Sam's Club Plus membership on a customer's spending at Walmart.com.

---

[3]That is, you observe the customer's spend and their membership status but not their household income and size, nor any of the other determinants of how much they spend on Walmart.com. This is the reason why we use the letter u, it is mnemonic for *unobserved*.

(c) (2p) Is $E[u_i] = 0$ an assumption or a normalization? Show it.[4]

Solution: $E[u_i] = 0$ is the result of normalization - we can simply define the terms in the model, including u, in such a way that $E[u_i]$ is made to be zero and the model holds. Suppose for the sake of argument that we find that $E[u_i] \neq 0$. However, we can just manipulate the terms as follows:

$$y_i = \alpha + \rho D_i + u_i - E[u_i] + E[u_i]$$

$$y_i = (\alpha + E[u_i]) + \rho D_i + (u_i - E[u_i])$$

Define $\tilde{u}_i = (u_i - E[u_i])$, and $\tilde{\alpha} = (\alpha + E[u_i])$, giving us a new expression for y_i:

$$y_i = \tilde{\alpha} + \rho D_i + \tilde{u}_i$$

and we have arrived at the same model as before, where $E[\tilde{u}_i] = 0$.

In general, if we find that for a given α and ρ, $E[u_i] \neq 0$, we can normalize by absorbing the non-zero mean of the error term into the constant term α, thus giving u_i the property that $E[\tilde{u}_i] = 0$.

(d) (2p) Let $(\widehat{\alpha}, \widehat{\rho})$ denote the Ordinary Least Squares (henceforth OLS) estimator of parameters (α, ρ) in model (1). Do you need to make any assumption on u_i to compute $(\widehat{\alpha}, \widehat{\rho})$ in a particular sample?

Solution: To compute the Ordinary Least Squares (OLS) estimators $(\widehat{\alpha}, \widehat{\rho})$ for the parameters (α, ρ) in the model $y_i = \alpha + \rho D_i + u_i$, you don't need to make any specific assumptions about the distribution or properties of the error term u_i.

The OLS estimators for α and ρ are obtained by minimizing the sum of squared differences between the observed values of y_i and the values predicted by the model for a given sample:

$$\widehat{\alpha}, \widehat{\rho} = \arg\min_{\alpha, \rho} \sum_{i=1}^{n} (y_i - \alpha - \rho D_i)^2$$

This estimation method does not require any assumptions specifically about the distribution or properties of u_i.

However, while OLS doesn't require explicit assumptions about u_i for estimation, certain assumptions under the classical linear regression are needed to establish the properties of the estimators (e.g., unbiasedness, consistency, efficiency). These assumptions include things like the error term having a mean of zero conditional on the predictors, homoscedasticity (constant variance of errors), and no correlation between the errors and the independent variables. Violations of these assumptions can affect the reliability of the OLS estimators in terms of their statistical properties.

---

[4]An assumption imposes a restriction on the objects/items present in your model, the restriction may or may not hold. For example, if a claim is stated subject to an assumption then your proof will use the assumption to arrive at the result, which means that the result may not obtain had you dropped the assumption. A normalization is when you recognize that two (or more) objects/items in a model are not separately identified, i.e., there is no way to learn about each of them separately, e.g., you may only learn their sum or product, or some other function of the two (or more) objects. If you recognize such a situation in your model you reparametrize the model so that the objects in the reformulated model are learnable.

(e) (10p) Verify expression (2). **Hint**: Leverage the derivations you did for PSet1, do not use linear algebra. **Note**: We leverage this result repeatedly, make sure to understand it both from an intuitive standpoint and algebraically.

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\rho} \end{bmatrix} = \begin{bmatrix} \bar{y}^0 \\ \bar{y}^1 - \bar{y}^0 \end{bmatrix}. \tag{2}$$

Solution: Given the OLS model described in part (d) we take first order conditions, obtaining a system of two equations with unknowns $(\hat{\alpha}, \hat{\beta})$:

$$\begin{cases} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\rho} D_i) D_i = 0 \\ \\ \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\rho} D_i) = 0 \end{cases}$$

Letting $\bar{y} \equiv \frac{\sum y_i}{n}$ (the sample average of spending) and $\bar{D} \equiv \frac{\sum D_i}{n}$ (the sample proportion of memberships), the solution to this system is

$$\hat{\alpha} = \bar{y} - \hat{\rho} \bar{D}$$
$$\hat{\rho} = \frac{\sum_{i=1}^n (y_i - \bar{y})(D_i - \bar{D})}{\sum_{i=1}^n (D_i - \bar{D})^2}$$

First let us examine $\hat{\rho}$. Let n_0 refer to the number of samples where $D_i = 0$ and n_1 refer to the number of samples where $D_i = 1$ (this necessitates that $n_0 + n_1 = n$). Also let $\sum_{D_i=1} y_i$ refer to the summation of sample expenditure for all members, and vice versa for $D_i = 0$. Then we get that

$$\hat{\rho} = \frac{\sum_{i=1}^n (y_i - \bar{y})(D_i - \bar{D})}{\sum_{i=1}^n (D_i - \bar{D})^2}$$

$$= \frac{\sum_{i=1}^n y_i (D_i - \bar{D})}{\sum_{i=1}^n (D_i - \bar{D})^2} = \frac{\sum_{D_i=0} y_i (0 - \frac{n_1}{n}) + \sum_{D_i=1} y_i (1 - \frac{n_1}{n})}{\sum_{D_i=0} (0 - \frac{n_1}{n})^2 + \sum_{D_i=1} (1 - \frac{n_1}{n})^2}$$

$$= \frac{\sum_{D_i=0} y_i (-\frac{n_1}{n}) + \sum_{D_i=1} y_i (\frac{n_0}{n})}{\sum_{D_i=0} (-\frac{n_1}{n})^2 + \sum_{D_i=1} (\frac{n_0}{n})^2}$$

$$= \frac{\frac{1}{n}(-n_1 \sum_{D_i=0} y_i + n_0 \sum_{D_i=1} y_i)}{\frac{n_0 n_1}{n^2} \cdot (n_0 + n_1)} = \frac{\frac{1}{n}(n_0 n_1)(-\frac{1}{n_0} \sum_{D_i=0} y_i + \frac{1}{n_1} \sum_{D_i=1} y_i)}{\frac{n_0 n_1}{n^2} \cdot (n_0 + n_1)}$$

$$= \frac{(-\frac{1}{n_0} \sum_{D_i=0} y_i + \frac{1}{n_1} \sum_{D_i=1} y_i)}{\frac{1}{n} \cdot (n_0 + n_1)} = \frac{(-\frac{1}{n_0} \sum_{D_i=0} y_i + \frac{1}{n_1} \sum_{D_i=1} y_i)}{n}$$

$$= \frac{1}{n_1} \sum_{D_i=1} y_i - \frac{1}{n_0} \sum_{D_i=0} y_i$$

$$= \bar{y}^1 - \bar{y}^0$$

Which is exactly what we want. Now let us show $\hat{\alpha} = \bar{y}^0$:

$$\begin{aligned}
\hat{\alpha} &= \bar{y} - \hat{\rho}\bar{D} = \bar{y} - (\bar{y}^1 - \bar{y}^0)\bar{D} \\
&= \bar{y} - (\bar{y}^1 - \bar{y}^0)\frac{n_1}{n} \\
&= \left(\frac{n_0}{n}\bar{y}^0 + \frac{n_1}{n}\bar{y}^1\right) - \left(\frac{n_1}{n}\bar{y}^1 - \frac{n_1}{n}\bar{y}^0\right) \\
&= \frac{n_0}{n}\bar{y}^0 + \frac{n_1}{n}\bar{y}^0 = \bar{y}^0
\end{aligned}$$

finishing the exercise.

(f) (2p) Use expression (2) to describe in plain English estimator $(\hat{\alpha}, \hat{\rho})$.

Solution: The estimator $\hat{\alpha}$ is the mean spending on Walmart.com for sampled customers that are not *Sam's Club Plus* members. Whereas $\hat{\beta}$ is the difference in mean spending on Walmart.com between sampled customers that have membership in *Sam's Club Plus* and sampled customers that do not have membership.

(g) (2p) **Without further assumptions**: Are $(\hat{\alpha}, \hat{\rho})$ unbiased/consistent estimators of (α, ρ)? Explain (no proof).

Solution: Without further assumptions, we cannot say that $(\hat{\alpha}, \hat{\rho})$ are unbiased or consistent estimators of (α, ρ). If u_i is correlated with D_i, the estimators could be biased and the overall correlation may not be interpreted as purely a causal effect. Consistency also cannot be guaranteed because there may exist serial correlation or heteroskedasticity in the error terms without necessary assumptions. Unbiasedness also requires that $E[U|D] = 0$, which is not an assumption that we can necessarily make in the context of this problem. In addition, consistency requires that $E[Y^2] < \infty$ and that $E[D_i^4] < \infty$, $\forall i$. While this assumption seems reasonable at a glance, it is still not necessarily sound with only the information we currently have on hand regarding this model and the sample.

(h) (4p) You have a random sample (RS) of Walmart.com customers. In econometrics the assumption that $E[u_i|D_i = 1] = E[u_i|D_i = 0]$ is called the "zero conditional mean assumption" (ZCMA) because, once we make the normalization $E[u_i] = 0$, it writes as $E[u_i|D_i = 1] = E[u_i|D_i = 0] = E[u_i] = 0$. Describe the ZCMA in plain English.

Solution: The ZCMA states that on average, the unobserved determinants have the same effect on the control and treatment groups. Moreover, individual determinants may have differing effects (positive/negative), but on average they will cancel out.

(i) (8p) Show that ρ is identified if ZCMA holds. **Hint**: Express ρ *exclusively* as a function of population data moments (PDM), that is, features of the population distribution of (y_i, D_i).

Solution: Suppose the ZCMA holds.

First note that,

$$\begin{aligned}
y_i(0) &= \alpha + u_i \\
\implies \mathbb{E}[y_i(0)|D_i = 0] &= \mathbb{E}[\alpha + u_i|D_i = 0] \\
&= \mathbb{E}[\alpha|D_i = 0] + \mathbb{E}[u_i|D_i = 0], \text{ by linearity} \\
&= \mathbb{E}[\alpha|D_i = 0], \text{ by ZCMA}, \\
&= \alpha, \text{ because } \alpha \text{ is a constant.}
\end{aligned}$$

Moreover,

$$\begin{aligned}
y_i(1) &= \alpha + \rho + u_i \\
\implies \mathbb{E}[y_i(1)|D_i = 1] &= \mathbb{E}[\alpha + \rho + u_i|D_i = 1] \\
&= \mathbb{E}[\alpha|D_i = 1] + \mathbb{E}[\rho|D_i = 1] + \mathbb{E}[u_i|D_i = 0], \text{ by linearity} \\
&= \mathbb{E}[\alpha|D_i = 1] + \mathbb{E}[\rho|D_i = 1], \text{ by ZCMA}, \\
&= \alpha + \rho, \text{ because } \alpha, \rho \text{ are constants.}
\end{aligned}$$

Combining these two results we can show that ρ is identified as follows:

$$\rho = \mathbb{E}[y_i(1)|D_i = 1] - \alpha$$
$$= \mathbb{E}[y_i(1)|D_i = 1] - \mathbb{E}[y_i(0)|D_i = 0].$$

(j) (12p) Assume that ZCMA holds. Is estimator $\widehat{\rho}$ unbiased (6p)? Is it consistent (6p)? Prove it. **Hint**: Let $D = \{D_1, \ldots, D_n\}$. Show $E[\widehat{\rho}] = \rho \ \forall \rho$ and $\widehat{\rho} \overset{p}{\to} \rho \ \forall \rho$ starting with expression (2).

Solution: By similar logic to what we argued in **Q1i**,

$$y_i(0) = \alpha + u_i$$
$$\implies \mathbb{E}[y_i(0)] = \mathbb{E}[\alpha + u_i]$$
$$= \mathbb{E}[\alpha] + \mathbb{E}[u_i], \text{ by linearity}$$
$$= \mathbb{E}[\alpha], \text{ by normalization,}$$
$$= \alpha, \text{ because } \alpha \text{ is a constant.}$$

Moreover,

$$y_i(1) = \alpha + \rho + u_i$$
$$\implies \mathbb{E}[y_i(1)] = \mathbb{E}[\alpha + \rho + u_i]$$
$$= \mathbb{E}[\alpha] + \mathbb{E}[\rho] + \mathbb{E}[u_i], \text{ by linearity}$$
$$= \mathbb{E}[\alpha] + \mathbb{E}[\rho], \text{ by normalization,}$$
$$= \alpha + \rho, \text{ because } \alpha, \rho \text{ are constants.}$$

Now we show that $\widehat{\rho}$ is unbiased. Let n_1 be the sample size of those who have *Sam's Club Plus* membership. Let n_0 be the sample size of those who do not have *Sam's Club Plus* membership. Consider the following:

$$\widehat{\rho} = \bar{y}^1 - \bar{y}^0$$
$$\implies \mathbb{E}[\widehat{\rho}] = \mathbb{E}[\bar{y}^1 - \bar{y}^0]$$
$$= \mathbb{E}[\bar{y}^1] - \mathbb{E}[\bar{y}^0], \text{ by linearity,}$$
$$= \mathbb{E}\left[\frac{\sum_{D_i=1} y_i(1)}{n_1}\right] - \mathbb{E}\left[\frac{\sum_{D_i=0} y_i(0)}{n_0}\right], \text{ by definition of sample mean,}$$
$$= \frac{1}{n_1}\mathbb{E}\left[\sum_{D_i=1} y_i(1)\right] - \frac{1}{n_0}\mathbb{E}\left[\sum_{D_i=0} y_i(0)\right], \text{ because } n_0, n_1 \text{ are constants,}$$
$$= \frac{1}{n_1}\sum_{1}^{n_1} \mathbb{E}[y_i(1)] - \frac{1}{n_0}\sum_{1}^{n_0} \mathbb{E}[y_i(0)], \text{ by linearity of independence, and that } y_i \text{ is drawn iid,}$$
$$= \frac{1}{n_1}\sum_{1}^{n_1}(\alpha + \rho) - \frac{1}{n_0}\sum_{1}^{n_0}\alpha, \text{ by argument above,}$$
$$= \frac{1}{n_1} \cdot n_1 \cdot (\alpha + \rho) - \frac{1}{n_0} \cdot n_0 \cdot \alpha$$
$$= \rho \forall \rho, \text{ by simplification.}$$

Let's now show that $\hat{\rho}$ is consistent. First, we can show that $\bar{y}^1$ is consistent for $E[y_i(1)]$. By definition we have that $\bar{y}^1 = \frac{\sum_{D_i=1} y_i(1)}{n_1}$. Because the collection of y_i is drawn i.i.d, we can argue that

$$\frac{\sum_{D_i=1} y_i(1)}{n_1} \xrightarrow{p} E[y_i(1)], \text{ by the Weak Law of Large Numbers}$$

Similarly, we can argue that $\bar{y}^0$ is consistent for $E[y_i(0)]$. We have that $\bar{y}^0 = \frac{\sum_{D_i=0} y_i(0)}{n_0}$. By the same logic,

$$\frac{\sum_{D_i=0} y_i(0)}{n_0} \xrightarrow{p} E[y_i(0)], \text{ by the Weak Law of Large Numbers.}$$

We can apply the Weak Law of Large Numbers by leveraging the fact that we have a random sample of Amazon.com customers (from Q1h), such that (y_i, D_i, u_i) are drawn i.id from some distribution. As shown above, $E[y_i(1)] = \alpha + \rho$ and $E[y_i(0)] = \alpha$, so we can argue:

$$\bar{y}^1 \xrightarrow{p} \alpha + \rho$$

$$\bar{y}^0 \xrightarrow{p} \alpha$$

By the Continuous Mapping Theorem, we can conclude that

$$\hat{\rho} = \bar{y}^1 - \bar{y}^0 \xrightarrow{p} \alpha + \rho - \alpha \ = \rho \forall \rho$$

proving consistency as desired.

(k) (2p) In place of maintaining ZCMA, assume that $u_i \perp D_i$, where the symbol "$\perp$" signifies statistical independence. Does your answer to **Q1j** change? If it does, how?

> **Solution:** Our answer to **Q1j** does not change because independence implies mean independence. Furthermore, ZCMA alongside the normalization that $E[u_i] = 0$ is just to say that we have mean independence between D_i and u_i. Thus, assuming $u_i \perp D_i$ alongside the normalization of $E[u_i] = 0$ maintains ZCMA, and this independence assumption is sufficient to show that $\widehat{\rho}$ is an unbiased and consistent estimator of ρ.

(l) (2p) In light of your answers to the previous questions: Do you expect estimator $\widehat{\rho}$ to be unbiased/consistent when constructed using a sample of actual Walmart.com customers chosen at random? Explain.

> **Solution:** We do not expect estimator $\widehat{\rho}$ to be unbiased and consistent with a sample of actual Walmart.com customers because it is difficult to achieve ZCMA in this setup. Considering our answer to the first question, we can imagine there are certain unobservable variables that might have differing effect on the control and treatment groups. For example, income might be a factor in u_i and we can see that it might affect the treatment group more in the sense that customers from a lower income background might be more inclined to take advantage of the membership and therefore make more purchases.

Q2. (20p) Consider the time **before** Walmart introduced *Sam's Club Plus*. Sales leadership have come up with the idea of a *Sam's Club Plus* membership that offers cash back on all orders. Scientists want to design and carry out a randomized control trial (RCT) to estimate how different consumer spend would be on average with a *Sam's Club Plus* membership. The estimate would help stakeholders decide whether to roll-out a *Sam's Club Plus* program, and how to price it.

(a) (2p) Suggest two reasons why a customer's spend at Walmart.com may differ with versus without a *Sam's Club Plus* membership, all else the same.

> **Solution:** One reason that a customer's spend at Walmart.com might be higher with versus without a *Sam's Club Plus* membership is the savings that come with the membership. A customer might feel more inclined to purchase more items from Walmart due to the cash back offered with the membership. Another reason that a customer's spend at Walmart.com might be higher with versus without a membership is the increased loyalty that a customer might feel holding the membership. A customer with a membership might feel more attached to the brand and therefore make more purchases from Walmart.com.

(b) (12p) The Walmart scientists carried out an RCT: they randomly assigned (RA) *Sam's Club Plus* membership status to $10,000$ existing customers (at no charge) (treated group) and left the rest of the customers *as is* i.e., without the membership (control group). Provide a discussion of this RCT's possible limitations/challenges by following step by step the list of limitations given in `CAUS_intro.pdf`, that is, implementation hurdles, lack of generalizability, etc.

> **Solution:**
> - Implementation hurdles: This RCT might be expensive to carry out as the memberships are offered to a large group of customers at no cost, so the cash back benefits might decrease the company's revenue if many customers in the treated group choose to purchase more to take advantage of the cash back program.
> - Lack of generalizability: The sample chosen for the RCT must accurately represent the broader Walmart shopper demographic to ensure the results are generalizable. For example, changes in other pricing strategies or the availability of products during the RCT could also affect the spending behaviors of the treatment and control groups and therefore prevent us from generalizing the results to the population.
> - Attrition Bias: If members drop out of the RCT because they don't perceive immediate benefits from the Sam's Club Plus membership, the remaining participants may not accurately represent the average treated consumer.
> - Hawthorne Bias: Participants aware of their inclusion in the study may change their shopping behavior. For example, those with the membership might spend more to maximize perceived benefits especially knowing that they have the benefit now for free.

- Substitution Bias: Control group members might seek alternative ways to save money, such as using other discount services. This could lead to an underestimation of the Sam's Club Plus membership's effectiveness on spending behavior.
- Compliance Bias: This limitation appears less applicable to this case. Walmart's treatment is not the choice to purchase membership, but an assignment of the membership to customers directly. It seems unlikely that there is any way to "comply imperfectly" in a case like this, though it may be similar to the Hawthorne Bias. The possibility that a customer with membership reacts maliciously by intentionally not purchasing products from Walmart is one possible case, though unlikely. If it is true that customer's can purchase membership on their own whilst being part of the control group, such a situation would be more applicable.
- Contamination Bias: This is less likely to occur in the context of this experiment however there could be a situation where a non-member in the control group asks a member in the treatment group to perform proxy purchases on their behalf.

(c) (6p) The RCT is carried out. The post-experiment analysis sample includes the $10,000$ customers in the treatment group, and $10,000$ customers chosen randomly from the control group. For each sample customer, the data only records the customer's membership status and how much they spent at Walmart.com during the first month following the date of RA, i.e., $\{(y_i, D_i) \,|\, i = 1, \ldots, n = 20,000\}$. Can the scientists use the experimental variation in this data to estimate ρ in model (1) by OLS? Explain. How shall they interpret the resulting OLS estimate?

Solution: Should scientists use this sample to perform OLS estimator of the population ρ in model 1, experimental variation will cause different samples of the population to yield variation in the sample estimator $\hat{\rho}$. With a fixed $n = 10,000$, we believe that n is sufficiently large, such that the experimental variation resulting from this sample is minimal and will likely not cause our $\hat{\rho}$ to deviate much from the true population ρ value. However, we must interpret our OLS estimate in the context of our sample - i.e we must know that is not guaranteed to accurately reflect the true population value - rather, our estimator is just one of many values that might be obtained from performing this experiment. However, should we draw an infinite number of samples of size n and perform our experiment, the average of our estimates should approach the true population ρ, assuming that our estimators are unbiased and consistent.

In the context of the *Sam's Club Membership*, our sample OLS estimator $\hat{\rho}$ represents the approximate increase in spending on Walmart.com that can be attributed/is caused solely having the membership, with all other factors being equal.

Q3. (20p) The setting is as in **Q1** but for the following: y_i is determined by a customer's *Sam's Club Plus*-membership status (D_i) and other unobserved determinants (v_i) according to the heterogeneous treatment effects model:

$$y_i = \alpha + \rho_i D_i + v_i. \tag{3}$$

Note that ρ_i varies across customers. You have access to an RS of existing Walmart customers. As in **Q1**, the data is the collection $\{(y_i, D_i) \,|\, i = 1, \ldots, n\}$. Note that ρ_i is **not** included in your data, nor is v_i.

(a) (2p) Interpret ρ_i.

Solution: In the context of our new setup, ρ_i represents the individual-specific treatment effect for a given customer i. Unlike in Question (1), p_i varies for each individual customer, and thus the magnitude of the effect of a Sam's Club Plus membership on spending will vary per individual in the sample. Specifically, ρ_i captures the difference between the spending of an individual with Sam's Club Plus and the spending of the same individual without Sam's Club Plus.

(b) (4p) Think of each ρ_i as a draw from a distribution. Let $\rho \equiv E[\rho_i]$, $\rho_1 \equiv E[\rho_i | D_i = 1]$, and $\rho_0 = E[\rho_i | D_i = 0]$. Describe in plain English these three objects. Interpret in plain English the assumption $\rho = \rho_1 = \rho_0$. Speculate about why this assumption is called "no selection on gains" (specialize your answer to the situation being considered).

Solution: ρ represents the expected treatment effect across the entire population, this is the difference in spending that an individual with Sam's Club Plus and the spending of that same individual without Sam's Club Plus would have on average.

ρ_1 is the expected treatment effect across customers with Sam's Club Plus. This is the difference in spending that an individual with Sam's Club Plus and the spending of that same individual without Sam's Club Plus would have on average, when such individual is currently a member of Sam's Club Plus.

ρ_2 is the expected treatment effect across the customers without Sam's Club Plus. This is the difference in spending that an individual with Sam's Club Plus and the spending of that same individual without Sam's Club Plus would have on average, when such individual is currently not a member of Sam's Club Plus.

The assumption that $\rho = \rho_1 = \rho_0$ states the following: the difference in spending that an individual with Sam's Club Plus and the spending of that same individual without Sam's Club Plus would be on average the same, irrespective of current membership status. The effect on average is the same for the total population, the subset of the population that has membership, and the subset of the population that does not have membership. The assumption may be called "no selection on gains" because it highlights that the decision to obtain membership ("selection") is not influenced by the causal effect of potential gains in spending as a result of obtaining Sam's Club Plus. Selection is not determined by an individual's personally determined gain (in utility or other metrics) from the membership.

(c) (2p) Verify that you can rewrite model (3) as a simple linear regression model:

$$y_i = \alpha + \rho_1 D_i + u_i \text{ with } u_i \equiv v_i + (\rho_i - E[\rho_i|D_i = 1])D_i. \tag{4}$$

Solution:

$$
\begin{aligned}
y_i &= \alpha + \rho_i D_i + v_i \\
&= \alpha + \rho_i D_i + \rho_1 D_i - \rho_1 D_i + v_i \\
&= \alpha + \rho_1 D_i + \rho_i D_i - \rho_1 D_i + v_i \\
&= \alpha + \rho_1 D_i + (\rho_i - \rho_1)D_i + v_i \\
&= \alpha + \rho_1 D_i + (\rho_i - E[\rho_i|D_i = 1])D_i + v_i \\
&= \alpha + \rho_1 D_i + u_i \text{ where } u_i = v_i + (\rho_i - E[\rho_i|D_i = 1])D_i
\end{aligned}
$$

(d) (6p) Consider the OLS estimator of the slope parameter in model (4). In **Q1** you established that $\widehat{\rho_1} = \bar{y}^1 - \bar{y}^0$ and $\widehat{\rho_1}$ is unbiased for ρ_1 under the ZCMA $E[u_i|D_i = 1] = E[u_i|D_i = 0]$.

   i. What are the substantive benefits of this result? That is, what do you learn about the causal effect of the treatment?

   Solution: Establishing that $\widehat{\rho_1} = \bar{y}^1 - \bar{y}^0$ and that $\widehat{\rho_1}$ is unbiased for ρ_1 under the ZCMA allows us to conclude that even though ρ_i may vary for each individual treatment, our estimator $\widehat{\rho_1}$ still directly reflects the true average causal effect of a Sam's Club membership on Walmart.com spending for our population. That is to say, in determining the average treatment effect ρ_1, we can average out the individual differences in treatment effect - i.e find that $E[ITE] = E[\widehat{\rho_1}]$. Thus we gain insight into the causal effects of our treatment regardless of the heterogeneity of the treatment effect amongst individuals.

   ii. What does the ZCMA $E[u_i|D_i = 1] = E[u_i|D_i = 0]$ imply for the relationship between the unobserved (v_i) and observed (D_i) determinants of the outcome?

   Solution: The ZCMA $E[u_i|D_i = 1] = E[u_i|D_i = 0]$ implies that the unobserved and the observed determinants of the outcome are uncorrelated, in the sense that the unobserved determinants have the same overall effect on y_i regardless of the value of D_i.

(e) (6p) Under which additional condition does $\widehat{\rho}_1$ allow us to infer the average causal effect of treatment for the entire population, rather than only for the sub-population of *Sam's Club Plus* members?

> **Solution:** We want to ensure that the treatment effect $\widehat{\rho}_1$ estimated for *Sam's Club Plus* members using $\widehat{\rho}_1$ can be generalized to the entire population. Thus we need the potential outcome of an individual to depend only on its own treatment assignment and not on the treatment assignment of any other individual - i.e the spending of an individual with $D_i = 1$ is not influenced by the membership status of any other individual. Formally, this can be expressed as follows:
>
> $$y_i \perp D_j \forall \, (i,j) \text{ s.t } i \neq j$$
>
> Thus, regardless of which specific individuals are assigned the treatment, our analysis will arrive at the same overall outcomes, and thus can be generalized to the entire population, not just only to *Sam's Club Plus* members.

Q4. (10p) Take stock. You worked with two models: (1) $y_i = \alpha + \rho D_i + u_i$, see expression (1); and (2) $y_i = \alpha + \rho_i D_i + v_i$, see expression (3). You considered one estimator: the OLS estimator of the intercept and slope coefficients in a linear regression of customer spend on a constant term and an indicator of *Sam's Club Plus* membership status. You derived the assumptions that suffice for the OLS estimator of the slope coefficient to be unbiased for ρ in model (1), and to be unbiased for the mean of ρ_i in the sub-population of *Sam's Club Plus* members or in the entire population in model (3). What did you learn about interpreting the OLS estimator of the slope coefficient in a causal context? Write 3 to 5 sentences.

> **Solution:** For model (1), where the treatment effect is assumed to be homogeneous, unbiased estimation of the causal effect relies on the assumption of no unobserved confounding - that is, $E[u_i|D_i = 1] = E[u_i|D_i = 0]$. This assumption ensures that the treatment assignment is independent of potential outcomes after conditioning on observed covariates, allowing for causal interpretation of the OLS estimator.
>
> In model (3), unbiased estimation requires a different set of assumptions. The OLS estimator can be unbiased for the mean of ρ_i within the sub-population of *Sam's Club Plus* members specifically, and a different set of assumptions applies, i.e that $\rho = \rho_1 = \rho_2$. Furthermore, we were asked to consider the implications of heterogeneous treatment effect, and how we can derive the average treatment effect even for varying individual treatment effects under certain independence assumptions.