Predictive Analytics for Business Strategy

Project 1

McDermott

Final Project Outline (**Answers in bold**)

* **Name: Justin Zheng**
* **Data/topic: Apple – Bottom Line**
* **Reviewer:**

1. (10) Clearly state the treatment effect we are interested in.

**The effect of price on profitability. In the data, we have logPrice and logProfit.**

1. (10) What is the naïve model and what is the variable of interest?

**Variable of interest: logPrice**

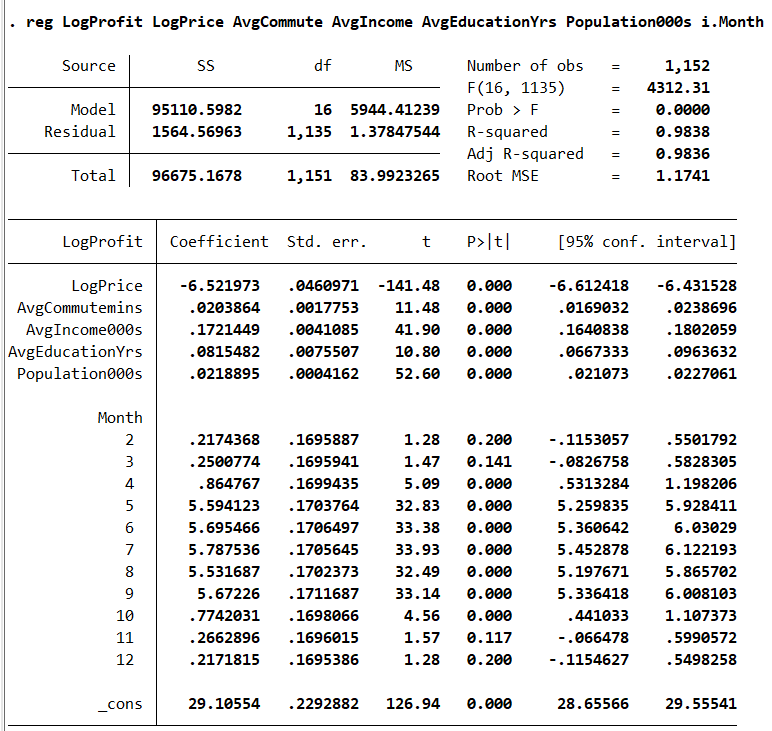
1. (5) Describe the endogeneity problem we have in the naïve model to explain why we can’t use the naïve model to get an unbiased estimate of the impact of the variable of interest on the outcome.

**The endogeneity problem is that some variation in log\_price is correlated with confounding factors in the U term. Confounding factors are variables that both affect the variation in the variable of interest and the target variable. For this project, it is stated that Apple is using regional pricing, meaning each record is for a different location. Thus, confounding factors might include demographics, average income, education, and more.**

1. (15) Using the four steps, identify confounding factors from the naïve model.
   1. **1. What is the treatment?   
      Treatment: prices of Apple products**
   2. **2. Who is Treated and Untreated:  
      Treated: Regions where prices are high  
      Untreated: Regions where prices are low**
   3. **3. Differences in Treated/Untreated**

* **Income: Regions with higher average annual incomes allow Apple to charge more, thus increasing the price in that region.**
* **Commute to work: Regions where the average person spends more time going to and from work mean the demand for phones is higher. A device linked to their car’s Bluetooth (or any other way safely used) will ease boredom. Thus, higher demand means higher price.**
* **Education: Regions where the average person’s education is more likely to buy Apple products, since they are more aware of the quality and features. Regions with higher education tend to have higher incomes. Both drives up demand. Thus, higher education means higher price.**
* **Population: Regions with more population means more potential customers, driving up demand. Thus higher populations means higher price.**
* **Month: Different months are likely to experience seasonality for Apple Products. Winter means Christmas, while Summer means vacation time. Those seasons might experience higher demand. Thus seasonality (month) can affect price.**
* **Age: Different age groups may have different preferences, thus can have different prices they are willing to pay.**
  1. **4. Which of these differences are in U?**
* **Income: Higher income positively and directly effects profits.**
* **Commute: Longer commute times positively and directly effects profits.**
* **Education: Regions with higher education years positively and directly effects profits.**
* **Population: Higher population count years positively and directly effects profits.**
* **Months: Differences in months directly effects profits.**
* **Average Age: DOES NOT directly affect profits since it does not affect (we don’t know for sure) overall demand.**

1. (15) State what the 3 categories are for confounding factors and categorize each confounding factor listed above into one of those three categories.
   1. **Factors that are fixed over time (for all entities)**
      1. **N/A**
   2. **Factors that change in parallel (for all entities)**
      1. **Months**
   3. **Factors that change differently for some entities.**
      1. **Income**
      2. **Commute**
      3. **Education**
      4. **Population**
2. (10) Type out your final (TWFE) model, show your the estimate of the model, and explain any key part(s) of the estimate. Assume it is exogenous (for this question) and make an active prediction.



**Assuming the model is exogenous, an increase in price by 1% would decrease profits by 6.52%**

1. (20) Explain how your TWFE model deals with the endogeneity issues that you have identified in the naïve model (as well as any you may not have identified). Be specific about how it handles the factors in each category and explain in the context/language of your specific problem.

**Factors that change in parallel**

**Create dummy variables for each period (month), and omit the first period to avoid perfect multicollinearity.**

**Factors that change differently for some entities**

**Often as confounding factors, we can add those variables in the TWFE model as controls.**

1. (15) In an earlier item, you attempted to have a model that was as close to exogenous as possible. Now, argue that the model you wrote is actually not exogenous, including anything you weren’t able to accomplish in the previous item as well as critiquing assumptions made in the previous item(s). Essentially, what are the weaknesses of your model in terms of exogeneity?

**Months may affect demand linearly, directly impacting price and profit. Month 1 may have the lowest demand, but Month 12 may have the highest. This means Month would no longer be a factor that changes in parallel but rather differently in some entities.   
Furthermore, there are possible factors that change differently for some entities but cannot be observed (becoming category 1 or 2) or for which we have no data.**