Problem Set 3

Econometrics I
NYU Stern, Fall 2018
Course Website:

Professor Paul T. Scott
Email: ptscott@stern.nyu.edu

Problem 1 Do Doctors Affect Drinking?

In this problem we will exploit the drinkData.csv dataset.[1] This data is taken from "The Effect Of Physician Advice On Alcohol Consumption: Count Regression With An Endogenous Treatment Effect," by Donald S. Kenkel and Joseph V. Terza (Journal of Applied Econometrics, 16: 165-184 (2001)). The goal of this paper was to understand if doctors could impact people's drinking activity. The authors do some sophisticated work to try and deal with concerns about causality. We will not replicate their methods. Instead, we will ignore issues related to omitted variable bias and focus on the tools of multiple regressions. There is a complete description of the dataset at the back of this problem set.

1. First let us get a feel for the dataset. The variable DRINKS is the number of drinks an individual has had and the variable ADVISE is a 0/1 variable for whether a person's doctor has told them to drink less. Report the mean number of drinks per person in each group. Similarly, calculate the mean education and income by group. Finally, what fraction of individuals in each group are between 30 and 40 and how many are between 40 and 50. Do a t-test for a dierence in group means of income and education. Do they appear to be different?

2. What is a possible source of omitted variable bias in a regression of DRINKS on ADVISE? You may think about the variables above or something else. Remember: omitted variable bias has two ingredients.

3. Regress DRINKS on ADVISE and do a one-sided significance test on ADVISE. What is the sign of ADVISE? Why do you think this might be the case?

4. Now run the same regression but including income, education and all the age dummies as controls. Report the results. What happens to the coefficient on ADVISE?

[1] Can be downloaded from ptscott.com/teaching/data/drinkData.csv.

5. Do an F-test to determine if age does not matter for drinking habits.

6. Create a variable called EVERDRINK that is a 0/1 variable for DRINKS being positive. Regress this on all age variables. Do an F-test to determine if the choice of whether to drink at all depends on age.

Problem 2 GPA and Parental Aid

Please read *More Is More or More Is Less? Parental Financial Investments during College* by Laura T. Hamilton http://asr.sagepub.com/content/early/2013/01/03/0003122412472680 which looks at the relationship between the GPA of college students and the amount of financial support they receive from their parents. We will on focus on this relationship (not the relationship between financial support and graduation).

1. Describe a potential causal mechanism by which parental financial support increases a student's GPA.

2. Describe a potential causal mechanism by which parental financial support decreases a student's GPA.

3. Describe the ideal experiment to measure the causal effect of parental contributions that the author is interested in.

4. Examine Table 3. Which regressors are continuous and which are categorical? How do we interpret the coefficient (in Model 2) of "Asian". Suppose we had made "Black" the omitted category, how would this change the estimated effect of "Asian"? How would the interpretation change? Would it also change the predicted value?

5. In Table 2, write the estimated regression equation for (Model 1). Be clear about the units (logs, levels, etc.). What is the interpretation of the coefficient on parental aid? Suppose an Asian Male has a GPA of 3.3 on a 4.0 scale. What happens to his GPA if his parents who were providing \$15,000 dollars now provide only \$10,000 of financial support?

6. As we move from (Model 1) to (Model 2) why might we worry that Parental Income is an important omitted variable?

7. Suppose we took the same student as before, whose parents earned \$50,000 per year and increased their income to \$55,000 per year and they provided all of the extra income in financial support to their son? Suppose they provided none of it in financial support to their son?

Problem 3 Supply, Demand and the Identication Problem

In this problem we will understand why economists began to think about instrumental variables in the first place. We will explore simultaneity bias. Suppose that you are an economist trying to predict the eect that raising tariffs will have on the price and output of coffee. To do this, you need to understand the supply and demand functions of coffee. For now, suppose we just want to understand demand. Suppose that demand has the following simple form:

$$Q^D = \beta_0 + \beta_1 P + \epsilon$$

This says that demand is linear in price but there are occasionally "demand shocks" that make everyone prefer more of a good or less of a good regardless of price. We suspect that $\beta_1 < 0$, as people demand less of a good when price is high.

Suppose that the market supply function for coffee has the following form:

$$Q^S = \alpha_0 + \alpha_1 P + \eta$$

We suspect now that $\alpha_1 > 0$ since as price is higher, the market is more willing to supply coffee since more producers will enter.

Finally, we know that in equilibrium, supply must equal demand. Now, you have collected data on many markets indexed by $m = 1, \ldots, M$. Each market has the same supply and demand functions but differ in their random shocks. We will work through the problem of estimating supply and demand. For simplicity, we will assume that $Var(\eta) = \sigma_\eta^2$, $Var(\epsilon) = \sigma_\epsilon^2$ and $Cov(\epsilon, \eta) = 0$.

1. First, think of at least one component of ϵ - something that will differ across markets that might affect demand for coffee that isn't the price of coffee. Do you think this would also affect supply?

2. Second, think of at least one component of η - something that will differ across coffee producers that might affect supply at any given price. Do you think this will also affect demand?

3. Now, use the fact that $Q^S = Q^D \equiv Q$ in equilibrium to solve for price and quantity in a market m as a function of η, ϵ and the model parameters.

4. Solve for $Cov(P_m, Q_m)$ in terms of σ_ϵ^2 and σ_η^2.

5. Solve for $Var(P_m)$ in terms of σ_ϵ^2 and σ_η^2.

6. A data scientist who hasn't thought too hard about the economics runs a regression of Q on P across markets and tells you that he has estimated demand. In other words, he believes that he has run this regression

$$Q = \beta_0 + \beta_1 P + \epsilon$$

From parts (4) and (5), solve for population version of the OLS estimator in this regression (i.e. $Cov(Q, P)/Var(P)$). Under what conditions, if any, will this be a consistent estimator of β_1.

7. Now suppose that we have data on whatever you suggested in part (2). Call this Z and assume that it is uncorrelated with ϵ, regardless of what you said: i.e. $Cov(Z, \epsilon) = 0$ and $Cov(Z, \eta) \neq 0$. Prove that if one uses Z as an instrument for price then, $\hat{\beta}_{1,IV} := Cov(Z, Q)/Cov(Z, P)$ is a consistent estimator of β_1.

8. The same logic as above allows us to use shocks to demand to estimate the supply function. Your friend the data scientist, wowed by your insights, suggests using the price of tea as a shock to demand for coffee. After all, if tea is expensive people will demand more coffee no matter the price. Do you think your friend has chosen a good instrument?

Variable	**Description**
DRINKS	Total drinks over a two week period
ADVISE	Dummy for being told to drink less by a doctor.
EDITINC	Monthly income ($1000)
AGE30	30 <age 40
AGE40	40 <age 50
AGE50	50 <age 60
AGE60	60 <age 70
AGEGT70	70 <age
EDUC	Years of schooling
BLACK	Black
OTHER	Non-white, non-black
MARRIED	Married
WIDOW	Widowed
DIVSEP	Divorced or separated
EMPLOYED	Employed
UNEMPLOY	Unemployed
NORTHE	Northeast
MIDWEST	Midwest
SOUTH	South
MEDICARE	Insurance through Medicare
MEDICAID	Insurance through Medicaid
CHAMPUS	Military insurance
HLTHINS	Health insurance
REGMED	Reg. source of care
DRI	See same doctor
MAIORLIM	Limits on major daily activ.
SOMELIM	Limits on some daily activ.
HVDIAB	Have diabetes
HHRTCOND	Have heart condition
HADSTROKE	Had stroke

Table 1: Variables in drinkData.csv