

# Problem Set 5

December 17, 2020

## Rules

- Solutions to problem sets must be completed and submitted individually.
- The deadline for submission is **23.59 on 8 January 2021**. Submissions after the deadline will not be accepted.
- The datasets for the STATA exercises can be found at SUCourse.
- **Submissions must include two files: 1) one word file including your answers (no hand-writing!!) 2) one clean STATA log file.** Results in the log file and interpreted results must match. Otherwise, you will get zero points for those exercises. **Do not upload zipped files.**
- You are expected to answer all questions very clearly. Therefore, please be careful about use of language and writing. **For instance; if you are asked to interpret a coefficient, you need to interpret it in a way that someone who cannot see the data and does not know econometrics can understand what you mean.**
- Double check that you write your name/surname and student ID number.
- Failure in fulfilling any of these will result in a FAIL grade for that homework.

## Question 1 - Price Elasticity of Demand for Air Travel (60 points)

Use the data in AIRFARE.DTA for this exercise. We are interested in estimating the price elasticity of air travel demand. The main variables in the data set are:

*passen*: average passengers per day.

*fare*: average airfare.

*dist*: the route distance (in miles).

1) (15 points) Estimate the price elasticity of air travel demand using **OLS**, controlling for the year specific effects (year dummies) and the route distance (*ldist* and *ldistsq*). What is the estimated price elasticity of air travel demand (Interpret the sign and magnitude of the coefficient)? Based on your estimation results what is the average change in demand for air travel from year 1997 to 2000?

2) (15 points) Estimate the same model in question (1) using **fixed effects (FE)** model. What is the estimated price elasticity of air travel demand? Explain why you are not able to estimate the effect of route distance on air travel demand with FE approach. Discuss which approach is preferable for the correct estimation of price elasticity parameter (FE or OLS)? Explain your answer.

3) (10 points) Since there might be a **simultaneity bias** in the estimation of price elasticity, we decide to use an instrumental variable (IV) approach. Consider the variable *concen*, which is a measure of market concentration. (Specifically, it is the share of business accounted for by the largest carrier.) Explain in words what we must assume to treat *concen* as a valid IV. Check whether *concen* satisfies the **relevance assumption**.

4) (20 points) Now, using the *concen* as an IV, estimate the demand function using **Two Stage Least Squares (2SLS)** estimator (you are expected to implement the two step estimation procedure as we discussed in the class.). What is the estimated price elasticity of demand? Next, estimate the same model using "ivreg" command in STATA. What is the main difference as compared to your previous 2SLS results?

## Question 2 - Education and Fertility (40 points)

The data in FERTIL2.DTA include, for women in Botswana during 1988, information on number of children, years of education, age, and religious and economic status variables.

1) (10 points) Estimate the model

$$children = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 age^2 + u,$$

by OLS, and interpret the estimates. In particular, holding age fixed, what is the estimated effect of another year of education on fertility? If 100 women receive another year of education, how many fewer children are they expected to have?

2) (10 points) The variable *frsthalf* is a dummy variable equal to one if the woman was born during the first six months of the year. Assuming that *frsthalf* is uncorrelated with the error term from part (1), show that *frsthalf* is a reasonable IV candidate for *educ*. (Hint: You need to do a regression.)

3) (10 points) Estimate the model from part (1) by using *frsthalf* as an IV for *educ*. Compare the estimated effect of education with the OLS estimate from part (1).

4) (10 points) Add the binary variables *electric*, *tv*, and *bicycle* to the model and assume these are exogenous. Estimate the equation by OLS and 2SLS and compare the estimated coefficients on *educ*. Interpret the coefficient on *tv* and explain why television ownership has a negative effect on fertility.