

# Final Exam

## ECON 301

January 12, 2021

### Rules

- Solutions to Final exam must be completed and submitted individually.
- The deadline for submission is **23.59 on 29 January 2021**. Submissions after the deadline will not be accepted.
- The datasets for the STATA exercises can be found attached at SUCourse-Final Exam.
- **Submissions must include two files: 1) one word file including your answers (no handwriting!!) 2) one clean STATA log file (do not include irrelevant analysis).** 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. **Quality of interpretations-discussions is going to be important in grading.** 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.
- The exam papers will be checked for originality on Turnitin. **Besides, some students might be interviewed in order to verify the originality of their exam papers.**
- 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.

## Price Elasticity of Fish Demand (55 points)

Use the data set in FISHEXAM.DTA, which comes from Graddy (1995). The data contains 97 daily price and quantity observations on fish prices at the Fulton Fish Market in New York City. Please check the definitions of the variables carefully and also use the "browse" command to see how they are recorded in the data. You are going to use this data to analyze the determinants of fish prices and to estimate a demand function for fish. In some of the questions, you need to generate new variables.

### Part 1: Determinants of Fish Prices

1) (10 points) Estimate an empirical model to analyze the determinants of fish price. The model needs to answer how price varies (**in percentage terms**) by different days of the week and over time (use quadratic time trend). Interpret and discuss your findings (coefficients, their significance and explanatory power of the model). Is there an evidence for a systematic variation in price within a week? What do the coefficients for quadratic time trend tell us?

2) (5 points) Now, add the variables "wave2" and "wave3" (to the above model), which are measures of wave heights over the past several days. Interpret the coefficients of these new variables. Are these variables individually significant? Explain why stormy seas would increase the price of fish. Explain why these variables can be assumed to be exogenous (not correlated with error term).

3) (5 points) Now, re-estimate the model in question (2) by using the **daily growth rate in fish price** as the dependent variable. Interpret the size of coefficients that are significant at 0.10 significance level. Is there a significant time trend? How can we explain the different results that we obtained for time trend in questions (2) and (3)?

### Part 2: Demand Function for Fish

4) (10 points) Now, you are expected to estimate the **price elasticity of fish demand**. Again, you need to control for daily seasonality and the quadratic time trend in your demand function. Discuss your findings. Interpret the size of the coefficients that are significant at 0.10 significance level. Discuss how your results might be affected when there is a random measurement error in "demand" variable? Discuss how your results might be affected when there is a random measurement error in "price" variable?

5) (10 points) The variables "wave2" and "wave3" are measures of ocean wave heights over the past several days. What assumptions do we need to make in order to use "wave2" and "wave3" as instrumental variables for fish price in estimating the demand equation? Discuss whether these assumptions are valid. Explain what

your results in question (2) indicate about validity of one of these assumptions.

6) (10 points) Now, estimate the model in (4) by 2SLS approach using "wave2" and "wave3" as instruments (here, you are expected to implement the two stage procedure in STATA). Next, estimate this 2SLS model with correct standard errors (using "ivreg" command). What is your conclusion about the price elasticity of fish demand? Based on this result, is the demand for fish price elastic or inelastic (check the definition of "elastic demand")? How can you explain the difference between elasticity estimates that are obtained in questions (4) and (6). What is the main methodological problem about the model estimated in (4) (Discuss the potential reason for a bias)? .

7) (5 points) Now, re-estimate the demand equation in question (6), this time by eliminating the outlier observations for "wave2" and "wave3" variables (do not include the days when the "wave2" or "wave3" are larger than 10). How did your price elasticity estimate change as compared to question (6)? What might be the other approach to eliminate the impact of these outliers on your result? (Hint: You can check the scatterplot showing the relationship between wave height and prices.)

### **Determinants of Crime Rate (35 points)**

Cornwell and Trumbull (1994) used data on 90 counties in North Carolina, for the years 1981 through 1987, to analyze crime rates. The data are contained in CRIMEEXAM.DTA. The crime rate is number of crimes per person, "prbarr" is the estimated probability of arrest, "prbconv" is the estimated probability of conviction (given an arrest), "prbpris" is the probability of serving time in prison (given a conviction), "avgsen" is the average sentence length served, and "polpc" is the number of police officers per capita.

8) (10 points) Discuss the summary statistics for the variables in the data. (For instance: What is the average crime rate in the counties etc.? Do this for all variables that you use in your model (in the next question))

9) (10 points) Estimate a model analyzing the determinants of crime rate in the counties. Include both crime related variables ( "prbarr", "prbconv", "prbpris", "avgsen" "polpc", "avgsen" ). and the control variables in your model. Here you are expected to build the best model, which will reduce the risk of bias for crime related variables. Interpret the coefficients that are significant at 0.10 significance level. Are the signs of these significant coefficients (especially crime related variables) in line with your expectations? If not, what might be the reason for surprising findings (explain clearly and provide an example reason for potential bias)?

10) (5 points) Now estimate the model in question (9) with Fixed effects Method. Discuss the main differences in your findings as compared to the results in question (9). What is the benefit of this method as compared to your estimation in question (9)? Explain. Why are some variables omitted from the regression? Explain.

11) (10 points) Your concern is that variable "polpc" is endogenous. What might be the reason for this concern? In order to deal with this problem, find an instrumental variable (IV) (from the ones available in the data) for "polpc" variable. Justify your choice. Discuss (if possible test) the validity of IV assumptions for the variable that you choose. Now, estimate the same model in question (9), this time with IV approach. Is there any difference in the estimated effect of "polpc" variable as compared to your finding in (9). Provide an example of a better IV candidate for "polpc" variable (if you could find data).

### Questioning the Reported Corona Death Numbers in Turkey (10 points)

12) (10 points) After the corona outbreak, there are some discussions about the accuracy of the corona-related death numbers that are reported by the governments. This is also a debate in Turkey. Some people argue that the number of corona-related deaths that are reported by the government is less than the real numbers. In this question, using city level monthly data, you are asked to build an empirical model (regression equation) that can help you test whether this argument is true or not. The variables that you are going to use are: **1)** official (reported by the government) number of **total** deaths for each city in Turkey for each month between March 2019 and June 2020 **2)** official (reported by the government) number of **corona-related** deaths for each city in Turkey for each month between March 2019 and June 2020 (this variable takes zero for the months before March 2020). For the governments, it is not easy to manipulate the total death numbers as it can be easily observed by the society, but it might be possible to manipulate the number of deaths that are related to corona (by changing the medical reports). Using these two variables and some additional time controls, describe a regression model that can help you test this argument. Which coefficient in this model is the coefficient of interest? If you want to test this argument, what does the null hypothesis for that coefficient needs to be?