Department of Economics
Columbia University

UN3412
Fall 2021

Make-up Final Exam
Introduction to Econometrics
(Erden)

Instructions:

First join the Zoom meeting at:
https://columbiauniversity.zoom.us/j/99677593629?pwd=Y1p6SkxqRHREUWlBNS9CQzgrN3ZvQT09

Meeting ID: 996 7759 3629

Passcode: 001498

You will upload your solutions to **Gradescope.**

This exam has six questions and an academic integrity statement that you have to "sign".

Please write your answers neatly by hand on paper, or type them into a Word document. If you choose to copy/paste some of your Stata output to Word best font is **`Bold Courier New Size 9`** for this purpose. When you are done, upload your Word document or scan and upload your handwritten solutions as a **single PDF** by merging all your answers. **Please upload your document to Gradescope when you are finished (just like you did with every problem set).** An extra 15 minutes has been provided for these tasks.

Make sure to specify the page number of each question when submitting to Gradescope. Once you upload your solutions to Gradescope you can exit the Zoom meeting.

Some questions ask you to draw a real-world judgment in a problem of practical importance. The quality of that judgment counts. For example, consider the question: "It is 10ºF outside. In your judgment, why are so many people wearing heavy coats?" The answer, "To stay warm" would receive more points than the answer, "Because they are fashion-conscious."

If you have any questions during the exam, you can ask questions through Zoom chat.

Question 1 [30 points]:

This question uses the data from the article “The Fulton Fish Market”.[1] The Fulton Fish Market was a colorful part of the New York City landscape that operated on Fulton Street in Manhattan for over 150 years. In 2005 the market moved from the South Street Seaport in lower Manhattan to Hunts Point in the South Bronx. The Fulton Fish Market--now called The New Fulton Fish Market--is one of the world's largest fish markets, second in size only to Tsukiji, the famous fish market in Tokyo. To economists, it may seem that a large centralized market with well-informed buyers and sellers should also be a very competitive market. But fish is a highly differentiated product. Buyers often wish to examine fish themselves, or have their agents do so. The centralized market performs an important function in matching fish to buyers. The table below describes the variables:

Variable name:	**Description:**
qty	Log of quantity of fish demanded
price	Log of average price of fish
day1	= 1 if observation belongs to Monday, = 0 otherwise
day2	= 1 if observation belongs to Tuesday, = 0 otherwise
day3	= 1 if observation belongs to Wednesday, = 0 otherwise
day4	= 1 if observation belongs to Thursday, = 0 otherwise
cold	= 1 if the weather is cold on shore, = 0 otherwise
rainy	= 1 if the weather is rainy on shore, = 0 otherwise
stormy	= 1 if the weather is stormy, = 0 otherwise
windspd	Wind speed
windspd2	Square of wind speed

[1] Graddy, Kathryn, “The Fulton Fish Market”, *Journal of Economic Perspectives*, Vol.20. No.2, pp 207-220, Spring 2006

(a) (4p) Using fishdata.dta fill out the following table (use the variable stormy as the instrumental variable in the 3rd and the 4th column:

	Dependent variable: log quantity			
	Ordinary Least Squares		Instrumental Variables	
	Column (1)	Column (2)	Column (3)	Column (4)
log price	( )	( )	( )	( )
monday	-	( )	-	( )
tuesday	-	( )	-	( )
wednesday	-	( )	-	( )
thursday	-	( )	-	( )
cold	-	( )	-	( )
rainy	-	( )	-	( )
constant	( )	( )	( )	( )
R^2				

(b) (3p) What is the price elasticity of demand of simple regressions under OLS? and under IV? (1st and 3rd columns)

(c) (3p) For the simple regressions in part (a), do 2SLS “by hand” with two separate regressions. Report the results of the first-stage and second-stage regressions, (remember when you copy/paste output from Stata **use font COURIER NEW font size 9 and bold** this way your output will look exactly as you see on the screen)

(d) (3p) Do you obtain the same coefficients and standard errors in part (c) as you did in part (a)? Why or why not?

(e) (3p) Outline the argument for the validity of the instrument *stormy* in words.

(f) (2p) Test the relevance of the instrument *stormy* for the regression in column (3) on the table.

(g) (3p) Consider adding additional instruments for *windspeed* and *windspeed squared*. Re-run the IV regression using all three instruments for both column (3) and column (4).

(h) (3p) Provide a test-statistic for the hypothesis that demand is unit elastic for all of the estimates in part (g)

(i) (3p) For the IV Estimates in part (g), provide a first stage F-test for a weak instrument. Do instruments appear to be weak? What does this mean for our demand elasticity estimate?

(j) (3p) For the IV estimates in column (4) provide a test of over-identifying restrictions (J-Test) and report its value. What does the J-test statistic indicate here?

Question 2 [5 points]:

What are the similarities and differences between ridge regression and forecasting?

Question 3 [5 points]:

The data set **macro_us.dta** contains quarterly data on real gdp (***rgdp***) and industrial production index (***ip)*** from 1960q1 to 2021q3.

(a) (2p) Run a autoregressive distributed lag (ADL(4,1) model for *rgdp* with four lags of *rgdp* and one lag of *ip*.

(b) (3p) Forecast rgdp for the 4th quarter of 2021 using the model in part (a)

Question 4 [8 points]:

Jose believes the following model is correct:

$$Wage = \beta_0 + \beta_1 * Schooling + \beta_2 * Experience + \beta_3 * Experience^2 + u$$

Robin thinks the following model is correct:

$$Wage = \beta_0 + \beta_2 * Experience + \beta_3 * Experience^2 + u$$

(a) (1p) Robin decides to follow Jose's model. If Robin's model is correct, will Jose's specification suffer from omitted variables bias?

(b) (1p) If Robin is right, what are the negative impacts of following Jose's specification?

(c) (1p) Camilla notes that you need a job to get a wage. Which one of the potential threats to internal validity does this relate to?

(d) (1p) Mark notes that if you have a high wage, you can pay for more schooling. Which one of the potential threats to internal validity does this relate to?

(e) (1p) If the range of wage increases with higher amount of schooling, what kind of problem would this cause in Jose's regression?

(f) (2p) John finds data for the same people over 5 years. Can he improve Jose's model? How? Explain.

<u>Question 5 [22 points]</u>:

Use data file named **airfare.dta** for this question. We are interested in estimating the model

$$\log(fare_{it}) = \vartheta_t + \beta_1 pconcen_{it} + \beta_2 \log\,(dist_i) + \beta_3 [\log(dist_i)]^2 + \alpha_i + u_{it}$$

where $fare$ is average one-way fare, $pconcen$ is the percent of market share of the largest airline (for given route) and $dist$ is distance in miles. Data is obtained from the Domestic Airline Fares Consumer Report by the U.S. Department of Transportation. ϑ_t means we allow for different year intercept. There are 1,149 routes and 4 years (1997-2000) in this data set.

Our major question here is whether higher concentration on a route causes higher airfares?

(a) (2p) Regress log of fares on $pconcen$. Report your results (remember when you copy/paste output from Stata **use font COURIER NEW font size 9 and bold** this way your output will look exactly as you see on the screen).

(b) (2p) Interpret the coefficient of $pconcen$

(c) (2p) Run the same regression as in part (a) but control for fixed effects (use xtreg) also make sure to use heteroskedasticity-autocorrelation-consistent errors.

(d) (2p) Now, consider above regression and interpret the coefficient of $pconcen$

(e) (2p) Why is your result in (a) and In (c) so different?

(f) (3p) Now run the same regression but this time controlling for entity fixed and time fixed effects (still make sure to use HAC errors)

(g) (2p) We added year dummies to the regression. How does your interpretation of $pconcen$ change?

(h) (2p) Do any of the models above control for distance? Why or why not?

(i) (2p) Now run the following regression and copy/paste your results ($passen_{it}$ is average passengers per day), interpret $\hat{\beta}_2$

$$\log(fare_{it}) = \beta_1 pconcen_{it} + \beta_2 \log\,(passen_{it}) + \alpha_i + u_{it}$$

(j) (3p) Now run the following regression controlling for airline fixed and time fixed effects and copy/paste your results, interpret $\hat{\beta}_2$

$$\log(fare_{it}) = \vartheta_t + \beta_1 pconcen_{it} + \beta_2 \log\,(passen_{it}) + \alpha_i + u_{it}$$

<u>Question 6 [30 points]</u>:

The data file **employment_08_09.dta** contains data on 5412 workers who were surveyed in the April 2008 Current Population Survey and reported that they were employed. The data file contains their employment status in April 2009, one year later, along with some additional variables.

Variable Names	**Description**
	Variables form the 2009 Survey
employed	= 1 if employed in 2009
unemployed	= 1 if unemployed in 2009
	Variables form the 2008 Survey
age	age
female	= 1 if female
married	= 1 if married
race	= 1 if self-identified race = white (only) = 2 if self-identified race = black (only) = 3 if self-identified race was not white (only) or black (only)
union	= 1 if a member of a union
ne_states	= 1 if from a northeastern state
so_states	= 1 if from a southern state
ce_states	= 1 if from a central state
we_states	= 1 if from a western state
educ_lths	= 1 if highest level of education is less than a high school graduate
educ_hs	= 1 if highest level of education is high school graduate
educ_somecol	= 1 if highest level of education is some college
educ_aa	= 1 if highest level of education is AA degree
educ_ba	= 1 if highest level of education is BA or BS degree
educ_adv	= 1 if highest level of education is advanced degree
earnwke	average weekly earnings
private	= 1 if employed in a private firm
government	= 1 if employed by the government
self	= 1 if self-employed

(a) Regress $employed$ on $married$ and $female$ and an interaction variable between $married$ and $female$. Copy/paste your regression output.

(i) (3p) Is there significant difference in probability of being employed for married males and married females? Write the null hypotheses and conclusion.

(ii) (3p) What is the predicted probability of a married female being employed? (Use two decimal points only and show your work)

(iii) (3p) What is the predicted probability of a married male being employed? (show your work)

(b) Regress $employed$ on age and $female$ and an interaction variable between age and $female$. Also test whether all your regressors are jointly significant. Copy/paste your regression output.

(i) (2p) Is the employment probability gap between gender widens or narrows as employers get older?

(ii) (3p) What is the predicted probability that a 30 year old female is employed, what is the predicted probability a 30 year old male is employed? (show your work)

(iii) (3p) What is the difference in predictive probability of a 40 year old and a 30 year old female being employed? (Make sure to write the units)

(c) (5p) Repeat part (b) using probit regression.

(d) (5p) Repeat part (b) using logit regression.