**Department of Economics** **UN3412**
**Columbia University** **Fall 2021**

**Problem Set 8**
**Introduction to Econometrics**
**(Seyhan Erden and Tamrat Gashaw)**

---

**Question #1 [Chapter 13]**

(30p) To analyze the effect of a minimum wage increase, a famous study used a quasi-experiment for two adjacent states: New Jersey and (Eastern) Pennsylvania. A $\hat{\beta}_1^{diffs-in-diffs}$ was calculated by comparing average employment changes per restaurant between the treatment group (New Jersey) and the control group (Pennsylvania). The difference-in-difference estimate $\hat{\beta}_1^{diffs-in-diffs}$ turned out to be 2.76 with a standard error of 1.36.

The authors also used a difference-in-differences estimator with additional regressors of the type

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1,t} + \ldots + \beta_{1+r} W_{r,i} + u_i$$

where $i = 1, \ldots, 410$. $X$ is a binary variable taking on the value one for the 331 observations in New Jersey. Since the authors looked at Burger King, KFC, Wendy's, and Roy Rogers fast food restaurants and the restaurant could be company owned, four $W$-variables were added.

(a) Given that there are four chains and the possibility of a company ownership, why did the authors not include five $W$-variables?

(b) OLS estimation resulted in $\hat{\beta}_1$ of 2.30 with a standard error of 1.20. Test for statistical significance and specify the alternative hypothesis.

(c) Why is this estimate different from the number calculated from $\Delta\overline{Y}^{treatment} - \Delta\overline{Y}^{control} = 2.76$? What is the advantage of employing this estimator of the simple difference-in-difference estimator?

(d) Let the vertical axis of a figure indicate the average employment fast food restaurants. There are two time periods, t = 1 and t = 2, where time period is measured on the horizontal axis. The following table presents average employment levels per restaurant for New Jersey (the treatment group) and Eastern Pennsylvania (the control group).

| | *PA* | *NJ* |
|---|---|---|
| *FTE Employment before* | 23.33 | 20.44 |
| *FTE Employment after* | 21.17 | 21.03 |

Enter the four points in the figure and label them $\overline{Y}^{treatment,before}$, $\overline{Y}^{treatment,after}$, $\overline{Y}^{control,before}$, and $\overline{Y}^{control,after}$. Connect the points. Finally calculate and indicate the value for $\hat{\beta}_1^{diffs-in-diffs}$

**Question #2 [Chapter 14]**

(40p) This question is based on an article authored by Hal R. Varian (2014) entitled "*Big Data: New Tricks for Econometrics*" published in The Journal of Economic Perspectives, Vol. 28, No. 2, pp. 3-27. Although you can download the article by searching it on CLIO, this article is posted together with this problem set. In his paper, the author describes a few of the tools for manipulating and analyzing big data. The author believes that these methods have a lot to offer and should be more widely known and used by economists.

You are required to read this article in its entirety and answer the following questions based on your reading.

a. (5p) Briefly summarize the article in your own words by focusing on the purpose of the paper and main message of the article.
b. (5p) What are the tools to manipulate Big Data?
c. (5p) What are the tools to analyze Big Data?
d. (5p) What are the general considerations for prediction?
e. (5p) What is a classification problem? How is it related to economists' explanatory variables or predictors?
f. (5p) In chapter 11, we used and discussed the HMDA data. On page 13, the author provides a data tree based on this data. Comment generally on the tree and his discussion in relation to our findings in chapter 11 (i.e., Table 11.2).
g. (5p) What is Boosting, Bagging, and Bootstrap (BBB)?
h. (5p) Discuss the variable selection section of the paper? What is the lasso and ridge regressions?

**Question #3 [Chapter 15 and 16]**

(60p) A model that attracted quite a bit of interest in macroeconomics in the 1970s was the St. Louis model. The underlying idea was to calculate fiscal and monetary policy impact and long run cumulative dynamic multipliers, by relating real output (growth) to real government expenditure (growth) and real money supply (growth). The assumption was that both government expenditures and the money supply were exogenous.

a. (6p) Visit the Federal Reserve Bank of St. Louis at https://fred.stlouisfed.org/ where you have access to the Fred Economic data and download the data for the required three variables (i.e., real GDP (GDPC1), real money supply (M2REAL), and real government expenditure (GCEC1)). The sample period should be from first quarter of 1960 to the fourth quarter of 2019 (i.e., 1960q1 – 2019q4). The real money supply (M2REAL) is available on a monthly frequency basis and don't forget to convert it into a quarterly frequency variable to match it with the other two variables. [*Hint*: Is M2REAL a stock or flow variable? Although, you may take the last month of each quarter or the three-month average as your quarterly value, here use the **first** month of each quarter.]

b. (6p) Import your data into Stata from the excel/csv files that you choose to download your data. You can also import these series directly from Fred data using a special Stata command (*freduse*) as shown https://www.youtube.com/watch?v=iiizhsX-I00. However,

the data frequency should be similar not to mix a monthly data with a quarterly data in the same Stata file. You can use this approach for GDPC1 and GCEC1 but you need to add M2REAL separately. Then create a *time* variable that has a quarterly format and let STATA know that *time* is the variable you want to indicate the data is time series. And compute (i.e., generate) growth rate of these three variables after you first transform them into natural logarithm and name/label them *ygrowth*, *mgrowth*, and *ggrowth*.

In order to investigate the effect of a fiscal and monetary policies on output, you want to estimate a St. Louis type model using your quarterly data (i.e., make sure to use HAC standard errors) and report your results. That is run a distributed lag model (DLM) using your dependent variable, (i.e., *ygrowth*) on:

c. (6p) Current period *mgrowth* and see the effect of a monetary policy on current quarter's output growth. Interpret the coefficient, make sure you are obtaining the correct standard errors. Is the effect significant?
d. (6p) Current period *ggrowth* and see the effect of a fiscal policy on current quarter's output growth. Interpret the coefficient, make sure you are obtaining the correct standard errors. Is the effect significant?
e. (6p) The effect of current and next quarter's monetary policy on output growth, making sure the standard errors are correct. What is the impact multiplier? Explain the meaning. What is the cumulative multiplier? Explain the meaning.
f. (6p) The effect of current and next quarter's fiscal policy on output growth, making sure the standard errors are correct. What is the impact multiplier? What is the cumulative multiplier?
g. (6p) The **change (i.e., first difference)** in current and four lags of *mgrowth* and *ggrowth* (i.e., to mimic the original St. Louis Equation) to obtain a regression like the one shown below (i.e., your regression coefficients should be different from this one) and report your results:

$$\widehat{ygrowth_t} = \underset{(0.004)}{0.018} + \underset{(0.079)}{0.006} \times dmgrowth_t + \underset{(0.091)}{0.235} \times dmgrowth_{t-1} + \underset{(0.087)}{0.344} \times dmgrowth_{t-2}$$

$$+ \underset{(0.097)}{0.385} \times dmgroth_{t-3} + \underset{(0.069)}{0.425} \times mgrowth_{t-4} + \underset{(0.049)}{0.170} \times dggrowth_t - \underset{(0.068)}{0.044} dggrowth_{t-1}$$

$$- \underset{(0.040)}{0.003} \times dggrowth_{t-2} - \underset{(0.051)}{0.079} \times dggrowth_{t-3} + \underset{(0.027)}{0.018} \times ggrowth_{t-4};$$

$R^2 = 0.346$, SER=0.03

h. (6p) Assuming that money and government expenditures are exogenous, what do the coefficients represent? Calculate the *h*-period cumulative dynamic multipliers from these. How can you test for the statistical significance of the cumulative dynamic multipliers and the long-run cumulative dynamic multiplier?

| Lag number | Monetary Dynamic Multiplier | Monetary Cumulative Multiplier | Fiscal Dynamic Multiplier | Fiscal Cumulative Multiplier |
|---|---|---|---|---|
| 0 | | | | |
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |

i. (6p) Sketch the estimated dynamic and cumulative dynamic fiscal and monetary multipliers (i.e., something similar to Fig 16.2 in your textbook).
j. (6p) For these coefficients to represent dynamic multipliers, the money supply and government expenditures must be exogenous variables. Explain why this is unlikely to be the case. As a result, what importance should you attach to the above results?

**Following questions will not be graded, they are for you to practice and will be discussed at the recitation:**

**1.** SW Exercise 14.1.
**2.** SW Exercise 14.2
**3.** SW Exercise 14.5.
**4.** SW Empirical Exercise 15.2.
**5.** SW Empirical Exercise 16.1.