**Week 2 Lab Handout- Time Series**

**PA 5033 – Multivariate Techniques**

March 25th, 2022

Wenchen Wang

**PART A: AD-HOC DISTRIBUTED LAG MODELS ~ 15min**

**PART B: DYNAMIC LAG MODELS~35min**

Use the data set on minimum wage in Puerto Rico (**PRtime data**). The data can be found on Canvas site for you to download.

This is time-series employment and wage data for Puerto Rico from 1950-1987.

**PART A: AD-HOC DISTRIBUTED LAG MODELS**

**1. Model #1: Standard Regression of Prepop as a Function of Kaitz and Yrn**

“**Prepop**t” is a variable that expresses the proportion of the Puerto Rican population that is employed in year t. “**Kaitz**t” is an index variable that expresses the Puerto Rican minimum wage in year t. The **kaitzt** index variable weights the minimum wage by its coverage of workers in Puerto Rico. “**Yrn**” is the year number in the sample. Run the regression and note the significance of the **kaitzt** variable.

***regress prepop kaitz yrn***

**

How do we interpret the coefficient on kaitz?

**2. Creating Lagged and Lead Variables for Kaitz**

First, because we will be running tests on time-series data, we must run the “tsset” command to tell STATA we are using time-series data and by which variable we are organizing the data using the command:

****

1. Lagged variables for Kaitz

* create a one-period lagged variable:

***gen kaitz\_1 = L.kaitz***

* create a two-period lagged variable:

***gen kaitz\_2 = L2.kaitz***

1. Lead variables for Kaitz

* create a one-period lead variable:

***gen kaitz\_3 = kaitz[\_n+1]***

* create a two-period lead variable:

***gen kaitz\_4 = kaitz[\_n+2]***

Now, you should have four new variables. **Kaitz**\_**1** is the value of **kaitz** lagged one period. **Kaitz**\_**2** is the value of **kaitz** lagged two periods. **Kaitz**\_**3** is the value of **kaitz** one year in the future. **Kaitz\_4** is the value of **kaitz** two years in the future.

**3. Model #2: prepop = f(kaitz, yrn**, **kaitz\_1, kaitz\_2, kaitz\_3, kaitz\_4**)

1. Run the regression as before but this time including all of the lag and lead variables for **kaitz**.



1. Notice the significance of the variables, particularly **kaitz**.

What happened to the coefficient on kaitz?

**4.** **Discussion**

What are the benefits and costs of using leads and lags?

Which model is preferable?

**PART B: DYNAMIC LAG MODELS**

1. **Run a regression with prunemp**, which is the Puerto Rican unemployment rate, as the dependent variable and **usgnp**, which is U.S. real gross national product, as the independent variable. This will measure the effect of the US GNP on Puerto Rican unemployment.



How do we interpret the coefficient for usgnp (GNP measured in billions)?

**Now create two new variables**:

Lagged variables for **usgnp**

* create a one-period lagged variable:

***gen usgnp\_1 = L.usgnp***

* create a two-period lagged variable:

***gen usgnp\_2 = L2.usgnp***

You will often need to create a new lag of the dependent variable also. In this case, you would create prunemp\_1 (Puerto Rican unemployment lagged one period) in the same way as above. However, this variable already exists in our data.

1. **Run the following regression: *prunemp = f(usgnp, usgnp\_1, usgnp\_2)***



Compare your results to those in part 2. How do you interpret the coefficients on usgnp\_1 and usgnp\_2? Why?

1. **Run the following regression: *prunemp = f(usgnp, prunemp\_1.)*** and request the Durbin-Watson statistic.



How do you interpret the coefficient on prunemp\_1?

What is the long-run multiplier?

How do you test for serial correlation in this model?

1. **Test for serial correlation in dynamic models:** DW statistics cannot be used to test for autocorrelation in dynamic models since our estimates are biased with serial correlation in dynamic models. We can use Lagrange Multiplier to do that with three steps:

Step 1: get residuals from the regression

*qui reg prunemp usgnp prunem\_1*

*predict e, resid*

Step 2: Create residuals lagged one period

*gen e\_1 = L.e*

Step 3: Regress residuals on explanatory & lagged residuals

*reg e usgnp prunem\_1 e\_1*

Step 4: Calculate LM statistics

LM = N \* R^2

**5. Test stationarity in dynamic models:** Dickey Fuller test:

Step 1: create first differences of USGNP

*gen usgnp\_d = d.usgnp*

Step 2: Run Dickey Fuller: regress difference on 1 lagged

*reg usgnp\_d usgnp\_1*

If we have non-stationarity, we can use first difference to try to solve this issue:

Step 1: Generate differences

*gen prunemp\_d = d.prunemp*

*gen prunemp\_1\_d = d.prunem\_1*

Step 2: Regress using differences, not levels

*reg prunemp\_d usgnp\_d prunemp\_1\_d*