

Homework 7
Due Thursday April 7 @ 1:30 PM
Full Credit will not be Given Unless You Show Your Work
Use Outreg to Display Regression Results

Part I - Computer Assignment – Go to the BEA website (bea.gov) and download quarterly observations on nominal (in \$billions) and real U.S. GDP (in billions of chained 2009 dollars) for 1947q1-2015q4. Then generate a Stata dataset. In your do file, include the following commands just after the “use” command

```
gen time=q(1947q1)+_n-1  
format time %tq  
tsset time
```

1. Display the summary statistics for nominal and real GDP using “tabstat” and interpret the mean values for these two variables.
2. Generate a graph of nominal and real GDP using the following command

```
twoway (scatter gdp time, ms(i) c(l)) (scatter gdpr time, ms(i) c(l))
```

use the “title” option to give the graph a title, use the “xtitle” and “ytitle” options to label the x and y axes and include the following as an option

```
legend(lab(1 "Nominal GDP") lab(2 "Real GDP"))
```

this option identifies the two curves in the graph as listed in the legend. Make sure to add a “,” after the “twoway” command above before listing the options. You can wrap lines using “///” for example

```
twoway (scatter gdpr time, ms(i) c(l)) (scatter gdp time, ms(i) c(l)), ///  
legend(lab(1 "Nominal GDP") lab(2 "Real GDP"))
```

is one command in Stata

3. Use a regression to calculate the average quarterly growth rates in nominal and real GDP. Make sure to state these growth rates. Display the regression results using “outreg”. Call this Table 1.
4. Estimate an AR model for real GDP. Follow the procedure in the class notes to determine the lag length (starting on page 7-49). Set $\alpha = 0.01$. You can use the “L.” notation in Stata to refer to lags. For example L.gdp and L2.gdp are the first and second lags of GDP. **Do this for observations through 2014q4.** For example

```
reg y x if time<=q(2014q4)
```

will limit the regression to this time period. Call this model GDPR(AR). Include the regression results for GDPR(AR) in Table 1. What is the optimal number of lags?

5. Out of sample forecasting: forecast real GDP for the 4 quarters 2015q1-2015q4. To do so re-estimate GDPR(AR) using the “arima” command (make sure to limit the sample to data through 2014q4). This allows for out-of-sample forecasting. To do so, after using the arima command type

```
predict ghatd, dynamic(q(2015q1))
```

The forecast error is the difference between the observed value of real GDP and its predicted value. Calculate the forecast error for real GDP for the periods 2015q1 through 2015q4. List these four forecast errors along with the time period. What is the mean absolute forecast error for these four quarters? How well do these forecasts fit the data?

6. Generate a graph of the observed and predicted values for real GDP for the period 2012q1 – 2015q4. Follow the directions in part 2 to generate this graph. Also include the following options

```
xline(220)
```

```
note("Note: the vertical red line indicates the beginning of the out-of-sample period")
```

Part II - Problems from the Text Book: 10.2, C10.2, C10.11

For problem C10.2, download the data set barium.txt from the class website. See Example 10.5 on page 357 in the text for background information on this model. The variables are (in order):

chnimp	Chinese imports, barium chloride
chempi	chemical production index
gas	gasoline production
rtwex	exchange rate index
befile6	= 1 for all 6 months before filing, = 0 otherwise
affile6	= 1 for all 6 months after filing, = 0 otherwise
afdec6	= 1 for all 6 months after decision, = 0 otherwise

Note that the data are monthly from February 1978 through December 1988. For problem C10.11, download the data set “traffic2.txt” from the class website. The data are monthly from January 1981 to December 1989. The variables are (in order):

year	1981 to 1989
totacc	statewide total accidents
unem	state unemployment rate
spdlaw	= 1 after 65 mph limit in effect, = 0 otherwise
beltlaw	= 1 after seatbelt law, = 0 otherwise
wkends	number of weekends in month
prcfat	percent of statewide total accidents that are fatal

Part III - Article: Read sections 3, 4, 5.1 and 5.2 of Binge Drinking and Labor Market Success: a Longitudinal Study of Young People,” by Shao-Hsun Keng and Wallace Huffman and answer the following questions

1. What are the two structural equations in the model? What are the endogenous and exogenous variables in each equation? What are the instruments that identify the system?
2. The authors are concerned that the price of alcohol variable in the binge drinking equation is endogenous? Why? How do the results change when the authors instrument for the price of alcohol?
3. The authors based their binge drinking variable on the “number of occasions when an individual consumed 6 or more drinks in one drinking session in the past 30 days.” But instead of using this measure, they make three sub-groups: 0 occasions, 1-3 occasions, and 4 or more occasions. Why do they do this?
4. The authors get very different results when they instrument for HEALTH? Why might HEALTH be endogenous in this model?
5. Do the results support the “bad rap” that college students have for binge drinking?
6. What is the impact of DRINKING on earnings? Is this impact economically significant?