Homework Set 3, ECO 311, Fall 2014

Due Date: At the beginning of class on December 4, 2014

Instruction: Each question is worth 2 points. You need to submit the answers of only five questions which you choose. The maximum point you can get is 10 points. For the purpose of preparing exam, you need to understand all questions. I will discuss the homework on the due date. Please do not ask me to go through the homework before the due date. However, you can discuss the homework with your classmates. You need to submit the homework individually though.

Q1: Two Sample T Test and Dummy Variable

Use the Wage data

1. (1.5 point) Report the results of following commands

\[
\begin{align*}
&\text{reg wage female} \\
&\text{sort female} \\
&\text{ttest wage, by(female)}
\end{align*}
\]

Interpret the coefficient of female produced by the command \textit{reg wage female}. Explain why the absolute value of t value for coefficient of female in command \textit{reg} is the same as the absolute value of two sample t test in command \textit{ttest}.

2. (0.5 point) From the \textit{ttest} table you may notice that the standard deviation is 4.160858 for male and 2.529363 for female. Can you modify the \textit{ttest} command in order to allow for unequal variances?

Q2: Interpreting Coefficients of Dummy Variables

Use Wage data

1. (1 point) Run the following regression \( lwage = \beta_0 + \beta_1 \text{female} + \beta_2 \text{married} + u \) where \( lwage \) is the log of wage. Please interpret \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \).

2. (1 point) Run a new regression \( lwage = c_0 + c_1 \text{female} + c_2 \text{married} + c_3 \text{female*married} + u \). Please interpret \( \hat{c}_3 \), and explain what the interaction term \textit{female*married} can capture.
Q3: Testing Difference in Regression Functions across Groups

Use Wage data. You may read pages 245-248 (5th edition) of the textbook

1. (1 point) Report two separate regressions, one for female and one for male. The regression is \( \log(wage) = \beta_0 + \beta_1 \text{educ} + u \). Please interpret the coefficient of educ in each regression.

2. (1 point) Test the null hypothesis at 5% level that the relation between log wage and educ does NOT depend on gender. Put differently, you need to run the Chow test for the null hypothesis of no structural change.

Q4: Equal and Unequal Weights

This problem intends to show the motivation for weighted least squares (WLS). Suppose a sample contains two independent observations \((X_1, X_2)\) with \(E(X_1) = \mu, \text{var}(X_1) = 4, E(X_2) = \mu, \text{var}(X_2) = 1, \text{cov}(X_1, X_2) = 0\). Note that the two observations have unequal variances (heteroskedasticity). There are two estimators for \( \mu \). The sample mean uses equal weights \( \bar{X} = \frac{1}{2}X_1 + \frac{1}{2}X_2 \). The weighted mean uses unequal weights \( \tilde{X} = \frac{1}{5}X_1 + \frac{4}{5}X_2 \).\( X_1 \) receives smaller weight because its variance is greater than \( X_2 \).

1. (1 point) Please show both \( \bar{X} \) and \( \tilde{X} \) are unbiased estimator for \( \mu \)

2. (1 point) Please show \( \tilde{X} \) is better (more efficient) than \( \bar{X} \) because \( \text{var}(\tilde{X}) < \text{var}(\bar{X}) \)

Q5: Heteroskedasticity-Robust Statistics and Weighted Least Squares

Please use Wage data

1. (1 point) report the conventional standard error, t value and p value and heteroskedasticity-robust standard error, t value and p value for the regression \( \text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u \). Compare them.

2. (1 point) report the WLS (feasible GLS) estimates for the above regression. Read page 286-288 of the textbook.

Q6: Working with Time and Date

Please use EZ data, where the variable \texttt{uclms} is the monthly unemployment claims in Anderson township in Indiana from January 1980 through November 1988. In 1984, an enterprize
zone (EZ) was located in Anderson. This data file is special because the time information is in two separate variables: the variable month is string (in uppercase), while the variable year is integer. We need to combine these two variables first.

1. (1 point) Please generate a string variable called time that combines month and year. Hint: the stata command is

   \[
   \text{gen time = string(year) + month} \\
   \text{gen timen = date(time, "YM")}
   \]

   We first transform the integer variable year to a string variable using function string, then combine it with month, which is string. Next we use function date to convert it to a time variable called timen, which is the days since January 1 1960.

2. (1 point) Please draw the time series plot for uclms. Hint: the stata command is

   \[
   \text{format timen %td} \\
   \text{tsset timen} \\
   \text{tsline uclms}
   \]

   Is uclms trending? Or showing seasonality?

**Q7: Monthly Dummy, Event Dummy, and Trend**

Please use EZ data, see Q7 for description of the data file.

1. (1 point) Please generate (in total 12) dummy variables, called monthly dummy, for January, February, ... and December. Denote them $D_1, D_2, \ldots, D_{12}$. Also generate the linear trend variable called trend. Hint: the command to generate January dummy and trend is

   \[
   \text{gen D1 = (month=="JAN")} \\
   \text{gen trend = [_n]}
   \]

   Note JAN are uppercase letters. Stata is case-sensitive. report the regression

   \[
   uclms = \beta_0 + \beta_1 D_1 + \ldots + \beta_{11} D_{11} + \beta_{12} \text{trend} + u.
   \]

   Please interpret $\beta_0, \beta_1, \beta_{12}$. Is there a trend in uclms? Note we use only 11 monthly dummies to avoid dummy variable trap.
2. (1 point) report the regression

\[ uclms = \beta_0 + \beta_1 D1 + \ldots + \beta_{11} D11 + \beta_{12} \text{trend} + \beta_{13} \text{ez} + u. \]

Please interpret \( \beta_{13} \).

**Q8: Autoregression and Forecasting**

This problem intends to show how to do forecasting using autoregressive (AR) model. Please use EZ data, see Q7 for description of the data file.

1. (1 point) report the AR(1) regression

\[ uclms_t = \beta_0 + \beta_1 uclms_{t-1} + u, \]

and AR(2) regression

\[ uclms_t = \beta_0 + \beta_1 uclms_{t-1} + \beta_2 uclms_{t-2} + u. \]

Which regression is better? Why?

2. (1 point) Please find the first-step forecast \( uclms_{n+1} \) and second-step forecast \( uclms_{n+2} \) based on the AR(2) regression, where \( n = 108 \).

**Q9: Philips Curve**

This problem intends to show how to collect data from internet, and verify the Phillips curve using DL and ADL models.

1. (1 point) From Fred data [http://research.stlouisfed.org/fred2/](http://research.stlouisfed.org/fred2/) download the monthly CPI and monthly civilian unemployment rate from January 1948 to August 2014. Put the two series into one excel file. Then read the excel file into stata, and generate the inflation rate by taking difference of log CPI

\[ \text{inflation}_t = \log(cpi)_t - \log(cpi)_{t-1}. \]

Draw the time series plots for the unemployment rate and inflation rate. Also draw the scatterplot of inflation rate against unemployment rate only using 1960-1969 data. Can you see the Phillips curve?
2. (1 point) Report the results of DL model and ADL model

\[ \text{inflation}_t = \beta_0 + \beta_1 \text{unemployment}_t + \beta_2 \text{unemployment}_{t-1} + u, \]

\[ \text{inflation}_t = \beta_0 + \beta_1 \text{unemployment}_t + \beta_2 \text{inflation}_{t-1} + u. \]

Please find the lag distributions from the two models. Which model makes more sense? Why?

Q10: Fisher Equation

In macroeconomics there is a theory called Fisher Equation, which relates normal interest rate \( i \) to expected inflation rate \( E(\pi) \) and real interest rate \( r \):

\[ i = E(\pi) + r \quad \text{(Fisher Equation)} \]

This problem asks you to obtain real interest rate using the Fisher Equation.

1. (1 point) From Fred data [http://research.stlouisfed.org/fred2/](http://research.stlouisfed.org/fred2/) download the monthly CPI and 3-Month Treasury Constant Maturity Rate (the webpage is [http://research.stlouisfed.org/fred2/series/DGS3MO](http://research.stlouisfed.org/fred2/series/DGS3MO)). We use the interest rate of Treasury Bill because it is risk-free. We use the 3 month rate because Fisher equation does not work well in short run. Put the two series into one excel file. Then read the excel file into stata, and generate the 3 month inflation rate by taking difference of log CPI

\[ 3 \text{ month inflation}_t = \log(cpi)_t - \log(cpi)_{t-3}. \]

2. (1 point) Suppose people use static expectation. That is, they let

\[ E(\pi) = 3 \text{ month inflation}_{t-1}. \]

Please generate the real interest and provide the time series plot of it. Please discuss how to obtain a smoother estimate of real interest rate. Also discuss how to modify the commands in order to get 1 year real interest rate.