Homework Set 3, ECO 311

Due Date: check syllabus

Instruction: There are 7 questions. Each question is worth 2 points. You need to submit the answers of only $\boxed{\text{five}}$ questions which you choose. The maximum point you can get is 10 points. For the purpose of preparing exam, you need to understand $\boxed{\text{all}}$ questions.

I will discuss the homework on the due date. Please <u>do not</u> 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: 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 $\texttt{var}(X_1) = 4, \texttt{var}(X_2) = 1$, $\texttt{cov}(X_1, X_2) = 0$. Note that the two observations have unequal variances (heteroskedasticity). There are two estimators for μ. 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. (2 point) Please show $\tilde{X}$ is better (more efficient) than $\bar{X}$ because $\texttt{var}(\tilde{X}) < \texttt{var}(\bar{X})$

Q2: Heteroskedasticity-Robust Statistics and Weighted Least Squares

Please use the 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 $\texttt{wage} = \beta_0 + \beta_1\texttt{educ} + \beta_2\texttt{exper} + u$.

2. (1 point) report the WLS estimates for the above regression. You may read section 8-4b of the textbook.

Q3: Monthly Dummy, Event Dummy, and Trend

Please use EZ data, where the variable `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.
(Hint) read data using these codes

```
ad = "https://www.fsb.miamioh.edu/lij14/311_ez.txt"
data = read.table(url(ad),sep = "\t", header=T)
attach(data)
head(data)
```

1. (1.5 point) Please generate (in total 12) dummy variables, called monthly dummy, for January, February,... and December. Denote them $D1, D2, \ldots, D12$. Also generate the linear trend variable called trend. Report the regression

$$\texttt{uclms} = \beta_0 + \beta_1\texttt{D1} + \ldots \beta_{11}\texttt{D11} + \beta_{12}\texttt{trend} + u.$$

   Please interpret $\beta_0, \beta_1, \beta_{12}$. Note we use only 11 monthly dummies to avoid dummy variable trap. (hint: check the variable "month" in the data; the trend variable should equal $1, 2, 3, ...n$)

2. (0.5 point) Report the regression

$$\texttt{uclms} = \beta_0 + \beta_1\texttt{D1} + \ldots \beta_{11}\texttt{D11} + \beta_{12}\texttt{trend} + \beta_{13}\texttt{ez} + u.$$

   Please interpret β_{13}. (hint: look at the data and see what ez is)

Q4: AR(2) Model and Forecasts

See last page of the slides for L7 (Time Series)

Q5: Statistical Significance and Economic Significance of Time Fixed Effects

Please use fatality data, and consider the following two-way fixed effect regression

$$y = \beta beertax + \alpha_i + \gamma_t + u_{it}$$

1. (1 point) test the statistical significance of time fixed effect (hint: use anova function)

2. (1 point) evaluate the economic significance of time fixed effect (hint: compare R squared)

Q6: Within Regression and plm function

Please use fatality data, and consider the following two-way fixed effect regression

$$y = \beta beertax + \alpha_i + \gamma_t + u_{it}$$

1. (1 point) use R function plm to report the estimation results
2. (1 point) use R coeftest function to report the cluster robust se, t value, and p value

Q7: First Difference Estimator with Panel Data

This problem does not require data. Consider the following one-way fixed effect regression

$$y_{it} = \beta x_{it} + \alpha_i + u_{it}$$

where α_i represents time-invariant unobserved factors that vary across i. In general, ignoring α_i leads to bias because

$$cov(x_{it}, \alpha_i) \neq 0.$$

Let us assume u_{it} causes no bias in the sense that

$$cov(x_{it}, u_{it}) = 0.$$

Please show me math to explain why it is a good idea to run a first difference regression in which the dependent variable is $\Delta y_{it} = y_{it} - y_{it-1}$ and the regressor is $\Delta x_{it} = x_{it} - x_{it-1}$.