Problem Set 2
Econometrics I

Vikrant Vig

Due: November 11th, 2016

1 Analytical Exercises

1. (15 pts) Suppose that

$$y_i = \beta y_{i-1} + e_i \tag{1.1}$$

where $E(e_i|y_{i-1}, y_{i-2}, ..., y_1) = 0$ and $E(e_i^2|y_{i-1}, y_{i-2}, ..., y_1) = \sigma^2$ for any integer $i > 1$. You know the values of $y_1, y_2, ..., y_N$ and you are asked to run OLS regression of $Y = (y_2, y_3, ..., y_N)'$ on $X = (y_1, y_2, ..., y_{N-1})'$. Do Gauss-Markov assumptions hold for you regression? What is $E(Y|X)$? What is $Var(Y|X)$? Do you think your estimate of β will be biased or unbiased?

2. (10 pts) Assume y is a $n\,x\,1$, X is $n\,x\,K$ (rank k), and $Z = XB$ where B is $k\,x\,k$ with rank k. Define $\hat{\beta}$ as the OLS estimator of y on X and $\hat{e}$ as the residuals of this regression. Defines $\tilde{\beta}$ as the OLS estimator of y on Z and $\tilde{e}$ as the residuals of this regression. What is the relationship between $\hat{\beta}$ and $\tilde{\beta}$, and $\hat{e}$ and $\tilde{e}$?

3. (15 pts) In class, we have seen the Gauss-Markov theorem. Under some assumptions, $Var(\hat{\beta}|X) \leq Var(\tilde{\beta}|X)$, where $\hat{\beta}$ is the OLS estimator and $\tilde{\beta}$ is any other linear and unbiased estimate. Does this result imply that $Var(\hat{\beta}) \leq Var(\tilde{\beta})$?

4. (10 pts) Suppose you have data on variables $y, x_1, ..., x_k$ all measured in dollars. Your coauthor has the same data on the same variables but measured in millions of dollars. Both of you are running the OLS regression y on a constant and $x_1, ..., x_k$. Consider the following parts of your stata printout: Coef, Std Err, t, R-squared. Which printout parts will be the same and which will be different?

2 Computational Exercises (Matlab and Stata)

5. (50 pts) For this exercise, you will need the dataset NLS_2006. There are 929 observations on nine variables in this dataset: **luwe** (log weekly age), **educ** (years of education), **exper** (years of experience), **age** (age in years), **fed** (father's education in years), **med** (mother's education in years), **kww** (a test score), **iq** (an iq score), and **white** (indicator for white).

   (a) Report the minimum, maximum, mean and standard deviation for all variables.

   (b) Estimate a linear regression model for log wages on education, experience, and experience squared. Report the estimates and standard errors. Also report the full variance/covariance matrix for all parameters, that is both the regression parameters as well as the residual variance.

   (c) Predict the effect on average log earnings of decreasing everybody's education level by one year. *(Hint: assume that exper = age - educ - 6)*

   (d) Can you obtain the above effect by running a regression with a redefined set of covariates? How?

   (e) (Partitioned Regression) Now regress log earnings and experience on education and experience squared. Compute the residuals of both regressions and then regress the residuals of the log earnings regression on the residuals of the exper regression. Discuss the coefficient and standard error estimates on this last regression. How do they compare with the ones in the regression of item (b)?