ECO 5350 Prof. Tom Fomby
Introductory Econometrics Fall 2016

TOPICS TO CONSIDER
FOR MID-TERM I

The first Mid-term exam in this class is scheduled for Thursday, October 6 at the regular class time but in *Room 241 Lee* where we will have more room to spread out. Don't spend a whole lot of time memorizing formulas. I will provide any formulas you need for the test. Of course, you should learn how to use the formulas. However, you should commit to memory the details of the basic ANOVA table that STATA produces when using the "regress" statement and what the table represents. Of course, you should know the basic decomposition of the sum of squares in multiple linear regression: SST = SSE + SSR and how the coefficient of determination (R-square) is calculated and interpreted. The orientation of the exam is more toward understanding how to use the tools of multiple regression for inference purposes and not so much toward theorem proving as in the unbiasedness of least squares estimates, the Gauss-Markov theorem, etc.

1. The test covers Chapters 1 – 4 in your Wooldridge textbook. Review your class notes and Exercises 1 – 5 and their keys that are posted on the class website. You have been given PPTs for the first four chapters. Study them closely. I have posted some lecture notes on the class website that I have used before in teaching this class. See, in particular, Lectures 1, 2, …, 14 in the "Lecture Notes" subdirectory. Print these notes out and study them. Also, I have given you the solutions to the odd-numbered problems (both "statement" and "computer" problems) for chapters 1 – 4 of the Wooldridge book. Look at them carefully. They should be very helpful.

2. What is the difference between experimental data and observational data? What do we mean by the term **causality**?

3. What are the different types of data sets that economists are called on to investigate? **Cross-Section data**, **Time Series data**, **Panel data**, and **Pooled Cross Section/Time Series Data**. What distinguishes these different data sets? Can you give me an example of each?

4. What does the term **homoskedasticity** mean? Draw a population regression function $(E(x \mid y))$ and several conditional density functions at various values of x while at the same time demonstrating the phenomenon of homoskedasticity. Do the same except draw a population regression and several conditional density functions at various values of x that demonstrate the phenomenon of **heteroskedasticity**. See, for example, Figures 2.1, 2.8, and 2.9 in your textbook. In looking at a data scatter of observations of y and x, how would you determine whether the population regression function is subject to homoskedasticity or heteroskedasticity? Is it OK to use OLS when you have heteroskedasticity in the errors of your regression model? If not, what should you use instead?

5. We have used STATA as our computer program of choice for this course. What is the difference between a .do file and a .dta file? What are the purposes of these files?

6. What are the **basic assumptions** of the simple regression model? Review assumptions SLR.1 – SLR.5 in your textbook. See chapter 2 in your textbook.

7. The ordinary least squares estimators of the coefficients β_1 and β_2 in the simple linear regression model are derived by using the calculus. What function do you minimize in order to get the OLS estimators? The nice thing about the least squares criterion is that it gives rise to **analytic solutions** for the coefficient estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ whereas when you use the **least absolute deviation criterion** there is no analytical formula for getting estimators, say $\tilde{\beta}_1$ and $\tilde{\beta}_2$. Can you write out the least absolute deviation criterion?

8. Study Table 2.3 in your textbook. Given different estimated equations, can you give an appropriate interpretation of β_1?

9. In the **sum of squares decomposition**, SST = SSE + SSR, what is SST, SSE, and SSR? What are their mathematical formulas?

10. What is an **SRF**? How is it different from a **PRF**? What is meant by the "**repeated sampling view**" of statistical hypothesis testing? In words, explain what is meant by "the **sampling distribution** of $\hat{\beta}_1$?" What is meant when we say "$\hat{\beta}_1$ is an **unbiased estimator** of β_1?" When we say $\hat{\beta}_1$ is a **consistent estimator** of β_1?"

11. What is the difference between Var($\hat{\beta}_1$) and $\widehat{Var}(\hat{\beta}_1)$?

12. Review the terminology in Table 3.1.

13. You should understand the **"partialling out" interpretation** of multiple regression, especially equation (3.22). What does $\hat{r}_{i1}$ stand for? So you can say that $\hat{\beta}_1$ is the covariance of ? with ? divided by the variance of ? You fill in the question marks in the previous sentence.

14. The stretch of pages 77 – 83 in your textbook (on specification analysis) is pretty important to understand. Also note the results reported in Table 3.2 on the bias in $\tilde{\beta}_1$ when x_2 is omitted in a multiple regression of y on x_1 and x_2. What happens to your least squares coefficient estimates when you **include irrelevant variables** in your regression? What happens to your least squares coefficient estimates when you **exclude a relevant variable** from your regression? How do you interpret the **omitted variable bias formula** $E(\tilde{\beta}) = \beta_1 + \beta_2\tilde{\delta}_1$?

15. What is the meaning of the term "multicollinearity?" How does it affect OLS inference?

16. What are the **properties of the least squares estimator** $\hat{\beta}_0$? What are the properties of the least squares estimator $\hat{\beta}_1$? What does it mean for these estimators to be **BLUE**? In other words, what does the **Gauss-Markov theorem** state? (See Section 3.5 in your textbook.)

17. Hypothesis Testing in Multiple Linear Regression – Chapter 4. Notice the additional assumption MLR.6 that has been added for this chapter. This assumption allows us to conduct small-sample tests of hypotheses. Notice the form of the t-ratio for testing the significance of variables in a multiple regression setting.

18. How do you display an estimated regression function (the SRF) when presenting it to someone else for his/her inspection?

19. How do you construct a **95% confidence interval for** β_1 using $\hat{\beta}_1$ and se($\hat{\beta}_1$)? How do you use such a confidence interval to test the null hypothesis that $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_1 : \beta_1 \neq 0$?

20. What is the difference between testing **a two-sided alternative** and testing **a one-sided alternative**? In a "word problem" would you be able to distinguish between when you would use a two-side test versus using a one-sided test? Can you get the p-value of an observed t-statistic for a two sided alternative? For a one-sided hypothesis? See Section 4.2 in your textbook.

21. What is the **coefficient of determination** (R^2)? How do you interpret $R^2 = 0.90$?

22. There are **three equivalent ways** to test a single linear combination of parameters in the linear regression model: (1) Direct computation of the t-statistic using the variance-covariance estimates of the coefficients in the regression model, (2) Reparametrization of the linear regression model to highlight the linear combination as a parameter on one of the variables of the reparametrized model and (3) Construct an F-statistic from the SSR's from an unrestricted versus restricted regression model and then take the square root of the F-statistic being careful to take the proper sign. How would you implement such techniques with a computer printout? See Section 4.4 in your textbook. You should practice on calculating the t-statistic for a test of the null hypothesis on a linear combination of coefficients in the multiple regression model. For examples see the discussion of the TWOYEAR data set, pp. 124 – 126.

23. How do you test multiple linear restrictions? The F-test with SSR_r and SSR_u. What kind of hypotheses does the **subset F-test** test? What are **exclusion restrictions**? See Section 4.5 in your textbook and equation (4.37). Also see the equivalent R^2 form in equation (4.41). Do you know how to calculate a p-value for an observed F-statistic? See, for example, Exercises 4.9 and 4.10.

24. What is the "**test of overall significance**?" What is the nature of the Analysis of Variance (**ANOVA**) Table in multiple regression analysis?

25. Do you know how to report an estimated multiple regression equation in **standard form**? See Section 4.6 in your textbook.

26. Just to let you know, I am really keen on our textbook because of the extensive use of examples in the textbook. I would encourage you to go through **each and every example** in chapters 1 – 4 of your textbook. They are there for a purpose – to help you build your intuition about estimation and hypothesis testing in multiple linear regression.