ECO 5350 Prof. Tom Fomby
Introductory Econometrics Fall 2007

TOPICS TO CONSIDER
FOR FINAL EXAM

The final exam is comprehensive. Therefore, you should review your previous exercises, QQs, mid-term review sheets, and mid-term exams as a starting point for your study. The topics we have covered since the second mid-term exam are listed below. See Sections XII, XIII, and XIV in your course outline under the topic heading of "Regression Analysis with Time Series Data" plus our wrap-up discussion of heteroskedasticity in cross-section data (tests of heteroskedasticity, WLS, and White's heteroskedasticity-consistent standard errors). I will provide a formula sheet for the exam so you need not go about memorizing a lot of formulas in preparation for the exam.

1. You should know the **consequences of heteroskedasticity** for the Ordinary Least Squares Estimator. See your textbook and Lecture Notes 20 for an extensive discussion of heteroskecastity, its consequences, and the ways we have of getting around the problems it causes.
2. You should understand how **White's heteroskedasticity-consistent standard error** is computed for the least squares estimator of the slope of a simple linear regression model. See my class notes and lecture notes on the website for the course. What are the **advantages** of using White's heteroskedastic-consistent standard errors and t-statistics in a multiple regression on cross-section data? What are the **disadvantages**? Do your coefficient estimates change when you adjust the standard errors and t-statistics? What method is used to get the coefficient estimates when using this **non-parametric approach** to handling the heteroskedasticity problem?
3. What **objective function** is minimized when using the **Weighted Least Squares** approach to estimating the coefficients of a multiple regression model? What is the **general formula for the weight** that one uses in Weighted Least Squares? (In EVIEWS the weight of $1/\sigma_i$ is applied to each observation on the dependent and explanatory variables including the intercept while in SAS they refer to the weight of $1/\sigma_i^2$ being applied to the i-th squared residual in the weighted sum of squared errors function.) If you assume **proportional heteroskedasticity**, what is the form of the weight that you use in applying the Weighted Least Squares method? (In Exercise 8 you use the proportional heteroskedasticity specification to estimate the regression equation there.) Be able to state **Aitken's theorem**. What is the **major implication** of Aitken's theorem? You should be able to graphically describe the theorem's result by drawing **competing sampling distributions** for the OLS estimator and the WLS estimator.
4. What is the form of **White's Heteroskedasticity test** with (and without) cross-product terms? What is the dependent variable of the test equation? What coefficients are tested in the test equation? What is the **null hypothesis of the test**? What is the **alternative hypothesis of the test**? If you accept the null hypothesis of the test what do you do next? If

you reject the null hypothesis what do you do next? Clearly state the **two options**. See Exercise 8 for White's test. What role do **residual plots** play in detecting heteroskedasticity? What residual plots should you always examine when estimating a cross-section regression equation? See Lecture 20 notes on the class website for more information on the topic of heteroskedasticity.

5. Write out what is meant by the autoregressive-distributed lag (**ARDL**) model. You need to know the process that is often used to "build" an ARX model. **First**, you build the autoregressive part of the model so that the last lag of the autoregressive part is statistically significant but, if you were to add one more autoregressive lag, it would not be statistically significant. Usually you build the model **bottom up** starting with one lag and then adding a second lag, etc. until the next lag that you add is statistically insignificant and thereby you have determined the lag order (p) of the autoregressive part of the model. **Second**, you can use the **cross-correlation function** of the pre-filtered data to get an idea of the distributed lag part of your model, that is, when the distributed lags start (i.e. with no delay, one-period delay, two-period delay, etc.) and the extent of the distributed lag, that is, its lag length. (What is the definition of the cross-correlation function? Be sure and know how to use it to get the delay and extent of the distributed lag parts of ARDL models.) **Third**, given the autoregressive core and a tentative distributed lag structure we can use the "top-down" approach to finalize the ARDL model for our data. **Fourth**, the residuals of our final model should be uncorrelated. In this case, we say that the ARDL model is "dynamically complete." See Exercise 9 for an example of this type of exercise. It might be noted that sometime the distributed lag part of the ARDL is difficult to determine because of the multicollinearity that almost certainly exists among the distributed lag variables. In this case we often use ALMON (polynomial) distributed lag structures to "smooth" the coefficients and enforce sensible lag structures on distributed lag coefficients. As I showed in class, the ALMON distributed lag specification imposes linear constraints on the coefficients of the ARDL model and these can be tested by a standard F-test. If the F-statistic of such a test is greater than 0.05 we can conclude that the distributed lag structure that we have imposed is sensible vis-à-vis the data. Otherwise, we have to rethink our chosen distributed lag shape. See **the first two pages of Time Series Regression Notes.doc** on the class website for more discussion of building a "dynamically complete" ARDL model.
6. The phenomenon of **spurious regression** and **unbalanced regression**. See my classroom lecture notes and look through the programs spurious.sas, spurious2.sas, and spurious3.sas and, in particular, the comment lines that I have put in to each program. What constitutes an unbalanced regression? What is likely to happen when running regression analyses on unbalanced variables? What might happen if you run a multiple regression involving two slow-turning variables?
7. When fitting a time series regression model we require that the variables, both dependent and explanatory, are transformed (if necessary) to stationary form. What does it mean to have a **stationary time series?** In terms of deciding on whether to difference data to make it stationary you should consult **Time Series Regression Notes.doc** and the discussion of **Augmented Dickey- Fuller Unit Root tests**. Do you know which of the three cases to apply to a given time series and how to read the results of a Unit Root Test from EVIEWS? What are the null and alternative hypotheses of the various cases? See the EVIEWS program unit_root_data.wf1. Graph the series. Test each of the series for a unit root. What are your

conclusions? Also see Lecture 24.pdf. Why are Dickey-Fuller Unit root tests a "must do" when doing multiple regression involving time series?

8. You need to come to understand how to interpret the **dynamics** of an ARDL model. See Lecture Notes 25 on the class website and go over the Dynamic.sas computer program that is posted on the class website. Also see Exercise 9 for an application of interim and total multipliers of an ARDL model to an advertising problem.
9. Topic of **Cointegration**. See Section 16.4 in your textbook and my classroom notes. Assume that you have two time series, X and Y, which are **cointegrated**. What two conditions must be met for the series to be cointegrated? In testing for cointegration you first must test the two time series for their stochastic orders by means of Augmented Dickey-Fuller tests. Second, you test the residuals of a regression of Y on X for a unit root. If the residuals exhibit a unit root then the null hypothesis of no cointegration is accepted. On the other hand, if the residuals appear to be stationary, we accept the alternative hypothesis that X and Y are cointegrated. Of course, we must remember that in EVIEWS the test of the residuals proceed by case 1 of the ADF test but we must get the p-values of the test from a table not available in EVIEWS (namely, the 1987 Engle/Granger paper or some other source).
10. Be able to give me at least three examples of economic/finance time series that have been shown to be cointegrated. What role does **arbitrage** play in the cointegration of time series?
11. Of course, cointegration, vis-à-vis **the Granger Representation Theorem**, implies that the two time series should follow an **Error Correction Model (ECM)**. Be able to write out an ECM for two variables. See Section 16.4 in your textbook. This is a "peculiar" equation system in that the equation has **both differences** of the variables (called the short-run dynamics of the model) as well **levels** of the variables entering in error correction terms (called the long-run equilibrium adjustment part of the model). What is an error correction term? What is an error correction coefficient? What do they represent? Of course, if there is cointegration, **at least one** of the error correction coefficients must be statistically significant. If both of the error correction coefficients are statistically significant, they should be opposite in sign, reflecting the equilibrium adjustment process.
12. **Sir Clive W.J. Granger** was a co-winner of the 2003 Nobel Prize in Economics. He was cited for his founding work on Co-integration. I am going to ask a few multiple-choice/fill-in-the-blank questions on the final exam that relate to the material he discusses in his Nobel Prize autobiographical sketch. Read his autobiographical sketch at http://nobelprize.org/nobel_prizes/economics/laureates/2003/granger-autobio.html