MFE230E Problem Set 2

Due April 7 10:00am via bCourses

You may NOT use built-in regressions routines for this problem set, i.e. construct the $\mathbf{Y}$ and $\mathbf{X}$ matrices and compute $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, standard errors, R^2, AIC/BIC and any other regression diagnostics yourself. You will receive NO credit if you use any regression package!

For the DF and ADF tests, your code should compute the DF and ADF test statistics but you can use the critical values of the DF and ADF tests from `statsmodels.tsa.stattools.adfuller`.

1. Assume that $x, \epsilon \sim i.i.d.N(0,1)$ and let $y = x + \epsilon$.

   (a) Consider the regression $y = \beta x + \epsilon$. Does this model satisfy the assumptions of the classical and/or asymptotic OLS models? Which OLS model is appropriate in this case? What are the theoretical properties of the OLS estimator β? What is the theoretical standard error of $\hat{\beta}$ for a sample size T?

   (b) Perform the following Monte Carlo simulation for $T = 20, 50, 100, 500$.

   Step 1: Draw T observations of x and ϵ and compute the implied y for each observation.

   Step 2: Compute the OLS regression $y = \beta x + \epsilon$.

   Step 3: Repeat steps 1. and 2. 10,000 times and save the $\hat{\beta}$ in each regression.

   Step 4: Plot the histogram of the distributions of the $\hat{\beta}$'s for each value of $T = 20$.

   Step 5: Compare the actual distributions of $\hat{\beta}$ from the simulations to the theoretical distributions.

2. Repeat question 1 assuming that $1 + x, 1 + \epsilon \sim i.i.d.\chi^2_1$. (Why do we add 1 to x and ϵ?)

3. Use the data in the spreadsheet `Tbill10yr.xls`. The file contains monthly yields of the 10-year Treasury bill. For this question, assume that the data is stationary.

   (a) Plot the data and perform a preliminary data analysis as discussed in class.

   (b) Plot the ACF and PACF. What do you learn from the ACF and PACF?

   (c) Compute the OLS regressions for the AR(1) and AR(2) models

$$x_t = \mu + \phi_1 x_{t-1} + \epsilon_t$$
$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$$

   (d) Report the OLS coefficients, their standard errors assuming homoskedasticity and heteroskedasticity, and other standard regression output.

   (e) Compute the roots of both models.

   (f) Plot the impulse response functions for the AR(1) and AR(2) models.

   (g) Compare the two AR models using the criterion that were discussed in class. Which model is preferred? Why?

4. Now, let's consider whether the T-bill data is stationary or not.

   (a) Use the methods discussed in class to test whether T-bill yields are stationary or not.

   (b) Unless you find overwhelming evidence of stationarity, consider alternative specifications in order to find a stationary model.

   (c) Consider various AR(p) models for the stationary specification. What is the "best" AR(p) model? Why? Try to be as comprehensive and thorough as possible.

   (d) Putting together everything you have learned about the T-bill data (including the results from the previous question), what is your preferred AR model for the T-bill data? Explain your choice.

5. Next, let's look at some spurious regressions. Repeat the following simulation exercise for $T = 200$ and $T = 1000$. For each T, simulate data with $\phi = 0.9$ as well as $\phi = 1$.

   (a) Simulate two independent AR(1) processes:

$$y_t = \phi y_{t-1} + \epsilon_t$$
$$x_t = \phi x_{t-1} + \eta_t$$

   where ϵ_t and η_t are two i.i.d. independent standard normal random variables.

   (b) Compute the OLS regression

$$y_t = \alpha + \beta\, x_t + u_t$$

   (c) Repeat (a) and (b) 10,000 times and save the OLS β, the t-test for $H_0 : \beta = 0$ and R^2 for each simulation.

   (d) Plot the histogram of the 10,000 βs, t-tests and R^2s.

   (e) Interpret the results. For example, describe how your results depend on T and ϕ.

6. Find a data series for a price level of some stock market index (e.g. DJIA, S&P, Russell or Wilshire indices) with at least 20 years of data (hint: Fred database at the St. Louis Fed at http://research.stlouisfed.org/fred2/ or Robert Shiller's webpage). Find some other data series on the internet that is arguably unrelated to the stock market price and illustrate a case of spurious regression, i.e. regress the log stock price on your variable and describe why your regressions is spurious. I will announce the three "weirdest" spurious regressions in class.