Problem Set

Individual Assignment, Due Friday 17 November

1. Consider the daily simple returns of the S&P 500 composite index from January1980 to December 2008. The index returns include dividend distributions. The data file is**S&P500WeekDays** which has 9 columns. The columns are (year, month, day, SP, M,T, W, H, F), where M, T, W, H, F denotes indicator variables for Monday to Friday,respectively. Use a regression model to study the effects of trading days on the index returns. What is the fitted model? Are the weekday effects significant in the returns at the 5% level? Use the HAC estimator of the covariance matrix to obtain the *t*-ratio of regression estimates. Does the HAC estimator change the conclusion of weekday effect?

2. The file **USMacro_Quarterly** contains quarterly data on several macroeconomic series for the United States: the data are described in the file **USMacro_Description**. Compute $Y_t = \ln(GDP_t)$, the logarithm of real GDP, and ΔY_t, the quarterly growth rate of GDP. In the problems below, use the sample period 1955:1-2004:4 (where data before 1955 may be used, as necessary, as initial values in regressions).

i. a) Estimate the mean of ΔY_t.

b) Express the mean growth rate in percentage points at an annual rate (*Hint*: Multiply the sample mean in (a) by 400.)

c) Estimate the standard deviation of ΔY_t. Express your answer in percentage points at an annual rate.

d) Estimate the first four autocorrelations of ΔY_t. What are the units of autocorrelations (quarterly rates of growth, percentage points at an annual rate, or no units at all)?

ii. Estimate an AR(1) model for ΔY_t. What is the estimated AR(1) coefficient? Is the coefficient statistically significantly different from zero? Construct a 95% confidence interval for the population AR(1) coefficient.

a) Estimate an AR(2) model for ΔY_t. Is the AR(2) coefficient statistically significantly different from zero? Is this model preferred to the AR(1) model?

b) Estimate AR(3) and AR(4) models. Using the estimated AR(1)-AR(4) models, use BIC to choose the number of lags in the AR model. How many lags does AIC choose?

ii. Use an augmented Dickey-Fuller statistic to test for a unit autoregressive root in the AR model for Y_t. As an alternative, suppose that Y_t is stationary around a deterministic trend.

3. There has been much talk recently about the convergence of inflation rates between many of the OECD economies. You want to see if there is evidence of this in North America by checking whether or not Canada's inflation rate and the United States' inflation rate are cointegrated.

   (a) You begin your numerical analysis by testing for a stochastic trend in the variables, usingan Augmented Dickey-Fuller test. The *t*-statistic for the coefficient of interest is asfollows:

Variable with lag of 1	*InfCan*	Δ*InfCan*	*InfUS*	Δ*InfUS*
t-statistic	–1.93	–5.24	–2.20	–4.31

   where *InfCan* is the Canadian inflation rate, and *InfUS* is the United States inflation rate.

   The estimated equation included an intercept. For each case make a decision about thestationarity of the variables based on the critical value of the Augmented Dickey-Fullertest statistic.

   (b) Your test for cointegration results in an Engle-Granger Augmented Dickey-Fuller (EG–ADF, see the lecture notes and Stock and Watson, 2007) statistic of (–7.34). Can you reject thenull hypothesis of a unit root for the residuals from the cointegrating regression?

   (c) Using a working hypothesis that the two inflation rates are cointegrated, describe how you would testwhether or not the cointegrating coefficient equals one.

   (d) Even if you could not reject the null hypothesis of a unit cointegrating coefficient, would that have beensufficient evidence to establish convergence?

4. In this exercise you will conduct a Monte Carlo experiment that studies spurious regression, a phenomenon where stochastic trends can lead two series to appear related when they are not.

   Generate two samples of *T*=100 i.i.d. standard normal random variables $\varepsilon_1, \dots, \varepsilon_{100}$ and $\eta_1, \dots, \eta_{100}$. (i) Set $Y_1 = \varepsilon_1, X_1 = \eta_1$, and $Y_t = Y_{t-1} + \varepsilon_t, X_t = X_{t-1} + \eta_t$, *t*=2, …, 100.

(ii) Regress Y_t onto a constant and X_t. Compute the OLS estimator, the regression R^2 and the *t*-statistic testing the null hypothesis that the coefficient β_1 on X_t is zero.

Use this simulation to answer the following questions.

(a) Run simulation (i) once. Use the *t*-statistic from (ii) to test the null hypothesis that $\beta_1 = 0$ using the usual 5% critical value of 1.96. What is the R^2 of the regression?

(b) Repeat (a) 1,000 times, saving each R^2 and the *t*-statistic. Construct a histogram of the R^2 and the *t*-statistic. What are the 5%, 50% and 95% percentiles of the distributions of the R^2 and the *t*-statistic? In what fraction of your 1,000 simulated data sets does the *t*-statistic exceed 1.96 in absolute value?

(c) Repeat (b) for different numbers of observations, for example, *T*=50, *T*=200 and *T*=500. As the sample size increases, does the fraction of times that you reject the null hypothesis approach 5%? Does this fraction seem to approach some other limit as *T* gets large? What is the limit?