JBS Master in Finance Econometrics Module

Lent 2010
Thilo Klein
http://thiloklein.de

Computer Lab Session 1
Unit Root Test and ADL Models

Contents

Exercise 1. Unit Roots

Use <u>real.csv</u>. The file contains the logarithms of the real exchange rates for Canada, Japan, Germany and the UK. Each series is constructed as $r_t = e_t + p_{ft} - p_t$ where

- r = log of the real exchange rate,
- e = log of the dollar price of foreign exchange,
- p_f= log of the foreign wholesale price index, and
- p = log of the US wholesale price index.

a)   For each sequence plot the data and find the ACF and PACF of **i)** the level of the real exchange rate; and **ii)** the first difference of the real exchange rate.

b)   Explain why is it not always possible to determine whether the sequence is stationary or non-stationary by the simple examination of the ACF and PACF.

c)   Including a constant, use an ADF (with trend, two lags) test to test whether the series are unit root processes.

    i.   A t-table indicates that at the 95% significance level, critical value is about 1.96 standard deviations. Why is it incorrect to conclude that the null hypothesis of a unit root can be rejected since the calculated t-statistic for $\gamma=0$ for RCAN series (with Trend+12 lags) is more than 1.96 standard deviations from zero?

    ii.   For each entry reported in the table what are the appropriate statistics to use (τ, τ_μ, or τ_τ) in order to test the null hypothesis of a unit root.

d)   Use the Japanese data to show that you can reject the null of two unit roots.

Answer (for RCAN):

    a) It is common practice when analyzing time series data to plot the autocorrelation and partial autocorrelation functions in order to try to guess the functional form of the data. To plot the autocorrelation and partial autocorrelation functions, use the ts library functions **acf()** and **pacf()**, respectively. The following commands plot the ACF and PACF on the same graph, one above (not on top of) the other.
> ➢ **par(mfrow=c(2,1))**
> ➢ **acf(RCAN)**
> ➢ **pacf(RCAN)**

    b) The ACF and PACF of a non-stationary process could be very similar to those of an ARMA model (ACF and PACF of ARMA processes decay slowly).

    c) ADF test.
> ➢ **RCAN_1 <- lag(RCAN, -1)**
> ➢ **RCAN_1.d1 <- diff(RCAN_1, lag=1)**
> ➢ **RCAN_1.d2 <- diff(RCAN_1, lag=2)**
> ➢ **trend <- ts(1:dim(real)[1], start=c(1973,1), frequency=12)**
> ➢ **adf.1c <- ts.union(RCAN.d1, RCAN_1, RCAN_1.d1, RCAN_1.d2, trend)**
> ➢ **lm1c <- lm(RCAN.d1 ~ RCAN_1 + RCAN_1.d1 + RCAN_1.d2 + trend, data=adf.1c); summary(lm1c)**
>
> **i)**   Because the t-statistics is not distributed t but DF, then it is wrong to use the t-distribution for critical values: **adf.test(RCAN, k=2)**
>
> **ii)**   For the first entry τ_μ, and for the last one τ_τ.

    d) DF test for Unit root in first difference:
> ➢ **adf.test(RJAP.d1, k=0)**
>
> We reject the null of unit root in the first difference.

Exercise 2. DL Models. *Orange juice price and freezing degree days*

Use <u>oj.csv</u>. The orange juice price data are the frozen orange juice component of processed foods and feeds group of the producer price index (PPI), collected by the U.S. Bureau of Labor Statistics. The orange juice price series was divided by the overall PPI for finished goods to adjust for general price inflation. The freezing degree days series was constructed from daily minimum temperatures recorded at Orlando area airports, obtained from the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce.

The dataset uses the following names for these series.

Variable Name	Description
ojfro	PPI for Orange Juice Frozen
ojfin	PPI for Finished Goods
frz	Freezing Degree Days

a)   Generate a time series of the % of change in the ojfro, call it cojfro. Plot the data. Test for unit roots.

b)   Estimate a static model, with cojfro as dependant variable and frz as independent variable.

c)   Estimate a DL model for this data with cojfro as dependant variable and frz as independent variable.

d)   Compute the impact and long-run multipliers

e)   Compute the cumulative multipliers (plot them including confidence intervals).

Answer:

   a)
   ojfro <- ts(oj$ojfro)
   cojfro <- ((ojfro/lag(ojfro,-1)) -1)*100
   plot(cojfro)
   Dickey-Fuller test:
   ➢ **adf.test(cojfro, k=0)**
   We reject the hypothesis of a unit root in this time series.

   b)
   frz <- ts(oj$frz)
   data2b <- ts.union(frz, cojfro)
   lm2b <- lm(cojfro ~ frz, data=data2b); summary(lm2b)

   c) Estimate the DL model using R's **dynlm()** function
   ➢ **library(dynlm)**
   ➢ **dl2c.18 <- dynlm(cojfro ~ L(frz, 0:18)); summary(dl2c.18)**
   There are many insignificant coefficients, let's simplify the model. Sequentially, eliminate the insignificant terms.

   d) The impact multiplier is 0.55 in the DL(12) and 0.51 in the DL(1). The long-run multipliers are 0.55 (0.55+0.14-0.14) and 0.64 (0.51+0.13) respectively.

   e) Take only the DL(1) model, assume that this is the "right model". The cumulative multipliers are: 0.51 ("impact") and 0.64 ("long-run"). To be able to estimate the standard deviations corresponding to each of these cumulative multipliers, let's estimate the following equivalent model (see appendix).

   $$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t$$

$$= \beta_0 + \beta_1 X_t - \beta_1 X_{t-1} + \beta_1 X_{t-1} + \beta_2 X_{t-1} + u_t$$
$$= \beta_0 + \beta_1(X_t - X_{t-1}) + (\beta_1 + \beta_2)X_{t-1} + u_t$$

or

$$Y_t = \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} + u_t$$

➢ **dl2e <- dynlm(cojfro ~ L(frz, 1) + d(frz, 1)); summary(dl2e)**

Let's plot the cumulative coefficients and coefficients +/- 2 std. error (95% confidence interval).

L.frz <- dl2e$coef[2]; d.frz <- dl2e$coef[3]
plot(c(d.frz, L.frz) ~ c(0,1), type="l", col="red", ylim=c(.4,.8))

se.L.frz <- summary(dl2e)$coef[2,2]; se.d.frz <- summary(dl2e)$coef[3,2]
lines(c(d.frz + 2*se.d.frz, L.frz + 2*se.L.frz) ~ c(0,1), col="blue")
lines(c(d.frz - 2*se.d.frz, L.frz - 2*se.L.frz) ~ c(0,1), col="blue")

Exercise 3: ADL model. El niño and fish population

Use soi.csv. The data shows monthly values of an environmental series called the Southern Oscillation Index (SOI) (the SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific) and associated Recruitment (number of new fish). Both series are for a period of 453 months ranging over the years 1950-1987. The central Pacific Ocean warms every three to seven years due to the El Niño effect, which has been blamed, in particular, for the 1997 floods in the midwestern portions of the U.S. Both series tend to exhibit repetitive behaviour, with regularly repeating cycles that are easily visible. The two series also tend to be somewhat related; it is easy to imagine that somehow the fish population is dependent on the SOI. Perhaps, even a lagged relation exists, with the SOI signalling changes in the fish population.

a)   Plot the data and test for unit roots.

b)   Estimate an Autoregressive Distributed Lag model.

c)   Compute the short- (or impact) and long-run multipliers

Answer:

a) Plot the data and test for unit roots.
   ➢ **soi <- ts(fish$soi, start=c(1950, 1), freq=12)**
   ➢ **rec <- ts(fish$rec, start=c(1950, 1), freq=12)**
   ➢ **par(mfrow=c(2,1)); plot(soi); plot(rec)**
   ➢ **adf.test(soi); adf.test(rec)**
   Both series are stationary.
b) An important advertence is that there is no "true" model. Beginning with an ADL(12,12) and after eliminating sequentially the less significant parameters we arrive to the following model.
   ➢ **adl3b.12.12 <- dynlm(rec ~ L(rec, 0:12) + L(soi, 0:12)); summary(adl3b.12.12)**
   ➢ **adl3b.12.8 <- dynlm(rec ~ L(rec, c(1,2,4:5,7,12)) + L(soi, c(5,6,8)))**
c) Note that there is no impact multiplier, in other words is equivalent to 0. The long-run multiplier is -83=(-21.5+8.1-3.3)/(1-1.22+0.37+0.12-0.19+0.09+0.03). This means that an increase in 1 point in the SOI reduces the recruitment by 82 in the long-run.

Appendix. Computation of cumulative multipliers and their standard errors

The cumulative multipliers can be computed by estimating the distributed lag model, then adding up the coefficients. However, you should also compute standard errors for the cumulative multipliers and while this can be done directly from the distributed lag model it requires some modifications.

One easy way to compute cumulative multipliers and standard errors of cumulative multipliers is to realize that cumulative multipliers are linear combinations of regression coefficients.

The trick is to rewrite the regression so that the coefficients in the rewritten regression are the coefficients of interest – here, the cumulative multipliers.

Example: Rewrite the distributed lag model with 1 lag:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t$$
$$= \beta_0 + \beta_1 X_t - \beta_1 X_{t-1} + \beta_1 X_{t-1} + \beta_2 X_{t-1} + u_t$$
$$= \beta_0 + \beta_1(X_t - X_{t-1}) + (\beta_1 + \beta_2)X_{t-1} + u_t$$

or

$$Y_t = \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2) X_{t-1} + u_t$$

So, let $W_{1t} = \Delta X_t$ and $W_{2t} = X_{t-1}$ and estimate the regression,

$$Y_t = \beta_0 + \delta_1 W_{1t} + \delta_2 W_{2t} + u_i$$

Then

$$\delta_1 = \beta_1 = \text{impact effect}$$

$$\delta_2 = \beta_1 + \beta_2 = \text{the first cumulative multiplier}$$

and the standard errors on δ_1 and δ_2 are the standard errors for the two cumulative multipliers.

In general, the ADL model can be rewritten as,

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \ldots + \delta_{q-1} \Delta X_{t-q+1} + \delta_q X_{t-q} + u_t$$

where

$$\delta_1 = \beta_1$$
$$\delta_2 = \beta_1 + \beta_2$$
$$\delta_3 = \beta_1 + \beta_2 + \beta_3$$
$$\ldots$$
$$\delta_q = \beta_1 + \beta_2 + \ldots + \beta_q$$

Cumulative multipliers and their SEs can be computed directly using this transformed regression.