Econ5121 B&C (Fall 2019)

Zhentao Shi

- Time Series Regression
- Univariate Time Series Models
- Nonstationary Time Series
- Multivariate Time Series

Useful R Packages

- `quantmod` : financial and US macro data
- `Quandl` : many data resources
- `dynlm` : single-equation dynamic model
- `tsDyn` : multiple-equation dynamic models

In [ ]:

```
library(quantmod, quietly = TRUE)
getSymbols("^HSI")

tail(HSI)
plot(HSI$HSI.Close, type = "l")
```

In [ ]:

```
library(Quandl)
CNH=Quandl("UNAE/GDPCD_CHN") #https://www.quandl.com/data/UNAE/GDPCD_CHN-GDP-Current-Prices-US-Dollars-China
HKG=Quandl("UNAE/GDPCD_HKG") #https://www.quandl.com/data/UNAE/GDPCD_HKG-GDP-Current-Prices-US-Dollars-China-Hong-Kong-SAR

head(CNH)
head(HKG)
```

Dynamic regression model

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \gamma y_{t-1} + e_t$$

Motivations

- temporal lags of effect. eg: policy lag
- expectation formed from the past. eg: forecast
- explicitly depends on history. eg: wealth accumulation

In [ ]:

```
library(quantmod)
getSymbols.FRED(Symbols = "POILBREUSDQ", env = .GlobalEnv) # Brent Oi
l price
x = POILBREUSDQ; T = length(x)
getSymbols.FRED(Symbols = "IPB50001SQ", env = .GlobalEnv) # Instrial
 Index (quaterly data)
Ty = length(IPB50001SQ); y = IPB50001SQ[(Ty - T + 1):Ty]

# quaterly data. Start 1990, end 2019

x = ts(x); y = ts(y)
plot(cbind(y,x), main = "")
```

ARDL(1,1) regression example

In [ ]:

```
library(dynlm)
reg = dynlm( y ~ L(y, c(1) ) + L(x,c(0:1) ) )
print(summary(reg))
```

Lagged Effect

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + e_t$$

Interpretation as a generative model

- Impact multiplier: β_0
- Cumulated effect (of τ periods): $\sum_{i=0}^{\tau} \beta_i$
- Equilibrium multiplier: $\sum_{i=0}^{\infty} \beta_i$

Lag Operator

$$Lx_t = x_{t-1}$$
$$L^\tau x_t = x_{t-\tau}$$

Difference operator

$$\Delta x_t = x_t - x_{t-1} = (1-L)x_t$$

Stationary time series

For a univariate time series $(y_t)_{t=-\infty}^{\infty}$,

- **Strictly stationary**: joint distribution of any finite coordinate only depends on their relative position.
- **Weakly stationary**: the first two moments of any pair y_t and y_s only depends on their relative position.
  - $E[y_t] = \mu$ for all t
  - $\mathrm{var}[y_t] = \sigma^2$ for all t
  - $\mathrm{cov}[y_t, y_{t+\tau}]$ only depends on τ independent of t

This notion can be extended to multiple-variate time series, for example (y_t, x_t, e_t).

Dynamic regression model

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + e_t = \alpha + B(L)x_t + e_t$$

where

$$B(L) = \sum_{i=0}^{\infty} \beta_i L^i$$

is a polynomial of the lag operators.

Autoregressive model

$$y_t = \alpha + \sum_{i=1}^{p} \gamma_p y_{t-p} + e_t$$

can be written as

$$C(L)y_t = \alpha + e_t$$

where

$$C(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p$$

is a polynomial of the lag operators.

Invertibility

If the roots of the polynomial equation $C(z) = 0$ **all** lies **outside** of the unit circle, we say the autoregressive model is invertible.

More generally, in the polynomial equation $C(z) = 0$, the root with the smallest module determines the trend of the time series.

If e_t is stationary with finite variance and $\alpha = 0$ (homogenous difference equation):

- If the module of the smallest root is bigger than 1, y_t is a stationary time series
- If the module of the smallest root is equal to 1, y_t is a **unit root** process
- If the module of the smallest root is smaller than 1, y_t is an **explosive** process

Numerical Example

- $C(L) = 1 - 0.5L$ is invertible.
- $C(L) = 1 - L$ is non-invertible.
- $C(L) = 1 - 1.1L$ is non-invertible.

In [ ]:

```
AR = function(b,T){
    y = rep(0,T)
    for (t in 1:T){
        if (t > 1) {
            y[t] = b * y[t - 1] + rnorm(1)
        }
    }
    return(ts(y) )
}
```

In [ ]:

```
T = 100; plot( x = 1:T, y = AR(0.5, T), type = "l")
```

In [ ]:

```
T = 100; plot( x = 1:T, y = AR(1.0, T), type = "l")
```

In [ ]:

```
T = 100; plot( x = 1:T, y = AR(1.05,T), type = "l")
```

Autoregressive Distributed Lag Models

ARDL(p,r) model:

$$C(L)y_t = \mu + B(L)x_t + e_t$$

where

$$C(L) = 1 - \gamma_1 L - \cdots - \gamma_p L^p$$

and

$$B(L) = \beta_0 + \beta_1 L + \cdots + \beta_r L^r.$$

Granger causality: $\beta_0 = \beta_1 = \cdots = \beta_r = 0.$

Model Specification

Information criterion.

Let k be the total number of slope coefficient in the model.

- Akaike information criterion: $\log(\hat{\sigma}^2) + 2 \times (k/T).$
  - Tend to overfit, but better for prediction
- Bayesian information criterion: $\log(\hat{\sigma}^2) + \log(T) \times (k/T)$
  - Model selection consistent

Seasonality

- Generated due to sampling frequency
- Add dummies to control seasonality

Spurious Regression

- The two time series $\{y_t\}$ and $\{x_t\}$ are generated independently, so that $E[y_t|x_t] = 0.$
- However, we observe a high R^2 and large t-value if we regression y_t against x_t.

In [ ]:

```
T = 50
a = 1

y <- AR(a, T)
x <- AR(a, T)
matplot( cbind(y, x), type = "l", ylab = "" )
```

In [ ]:

```
reg <- lm(y ~ x)
summary(reg)
```

Granger and Newbold (1974)

run a regression to check that if we naively use 1.96 as the critical value for the t-ratio, how often we would reject the null hypothesis that $\beta = 0$.

- The nominal asymptotic test size is 5% according to the standard asymptotic theory
- The empirical size is about 0.80 in this simulation
- The drastic deviation suggests that the standard asymptotic theory fails in the nonstationary environment.

In [ ]:

```
spurious <- function(i, a, T){
    y <- AR(a, T)
    x <- AR(a, T)

    reg <- lm(y ~ x)
    p.val <- summary(reg)[[4]][2,4]
    # save the p-value of the estimate of x's coefficient
    return(p.val)
}

library("plyr")
out <- ldply(.data = 1:1000, .fun = spurious, a = 1.0 , T = 100)
print( mean(out < 0.05) )
```

Error Correction Model

Cliver Granger (Nobel prize 2001)

Subtract y_{t-1} from both sides of the ARDL(1,1) model

$$\begin{aligned}\Delta y_t &= \mu + \beta_0 x_t + \beta_1 x_{t-1} + (\gamma_1 - 1) y_{t-1} + e_t \\ &= \mu + \beta_0 \Delta x_t + (\beta_1 + \beta_0) x_{t-1} + (\gamma_1 - 1) y_{t-1} + e_t \\ &= \mu + \beta_0 \Delta x_t + (\gamma_1 - 1)(y_{t-1} - \theta x_{t-1}) + e_t\end{aligned}$$

where $\theta = (\beta_1 + \beta_0)/(1 - \gamma_1)$.

- A short-run relationship $\Delta y_t = \mu + \beta_0 \Delta x_t + e_t$.
- An long-run equilibrium error $(\gamma_1 - 1)(y_{t-1} - \theta x_{t-1})$.

When y_t is nonstationary

- First difference recovers stationarity
- Useful to identify spurious regression
- Can be estimated either by OLS or by NLS