Project Description

To complete a project, you can choose a topic that interests you using your own model or follow the 5 steps below to identify a non-stationary series and use cointegration to find the correct recipe for a stationary combination! We also provided a demo after the instructions in case you don't know how to follow the steps listed.

- **Step 1 :** Get Data
- **Step 2 :** Are the series stationary? Unit Root Test to each variable
- **Step 3 :** Are the series cointegrated? Estimate the cointegrating parameter and do unit root test to the spread (linear combination)
- **(Step 4) :** Can we include the spread to estimate error correction model?
- **Step 5:** Plot the stationary spread series with bounds and discuss implications

Not sure what your project will look like? Check out this demo!

- **Step 1 :** Get Data

In this demo, we use data of stock prices for two stocks, namely "stocka" and "stockh", which can be directly fetched using functions in the R library quantmod. Try entering these scripts in your R Studio and the data will be at your hands!

```
library(quantmod);
# get H stock data for 1057.hk
setSymbolLookup(STOCKH=list(name="1057.hk",src="yahoo"))
getSymbols("STOCKH",from="2013-11-05",to="2016-11-05")
stockh <- log(STOCKH[,4])
# get A stock data for 002703.sz
setSymbolLookup(STOCKA=list(name='002703.sz',src='yahoo'))
getSymbols("STOCKA",from="2013-11-05",to="2016-11-05")
stocka <- log(STOCKA[,4])
```

- **Step 2 :** Are the series stationary? Unit Root Test to each variable.

Once we've got the price series for a pair of two stocks, we check if they are stationary separately using unit root test.

```
# unit root test for A and H stock
adf.test(stockh,k=1)
adf.test(stocka,k=1)
```

Now let's look at the output for one of the stock price series!

```
        Augmented Dickey-Fuller Test

data:  stocka
Dickey-Fuller = -2.0775, Lag order = 1, p-value = 0.5455
alternative hypothesis: stationary
```

The p-value shows the null hypothesis, i.e., the series of "stocka" being unstationary, is not a tail event (it's very likely to happen) so we do not reject it and accept that it's unstationary. The same rule applies to "stockh" and we do not repeat it here.

- **Step 3 :** Are the series cointegrated? Estimate the cointegrating parameter and do unit root test to the spread (linear combination)

Remember, we think that the "stocka" and "stockh" are cointegrated because they might have the same source of unstationarity, which can be eliminated by forming a linear combination of them. The ratio of this combination is actually the coefficient of one variable regressed on another. So let's run the regression!

```
# estimate cointegrating parameter
stocka.ts <- ts(stocka)
stockh.ts <- ts(stockh)
reg1 <- dynlm( stocka.ts ~ stockh.ts + L(d(stockh.ts)) + L(d(stockh.ts),2) + L(d(stockh.ts),-1) +
L(d(stockh.ts),-2))
beta <- reg1$coefficients[2]
```

The regression that the above code intends to run is:

$$\begin{aligned} stocka_t = \alpha + \beta stockh_t + \delta_1 \Delta stockh_{t-1} \\ + \delta_2 \Delta stockh_{t-2} + \gamma_1 \Delta stockh_{t+1} \\ + \gamma_2 \Delta stockh_{t+2} + e_t \end{aligned}$$

Notice here, beta is the coefficient that will be useful in forming the combination, simply compute a new series called "spread" and do unit root test to it.

$$spread_t = stocka_t - \beta stockh_t$$

Unit root test:

```
# unit root test to linear combination
spread <- stocka - as.numeric(beta) * stockh
adf.test(spread,k=1)
```

The rule to check the output will be the same as for "stocka" in the above example.

- **(Step 4) :** Can we include the spread to estimate error correction model?

This step is optional, interested readers can look up later chapters in the book to learn about error correction model. Here we provide an illustration for the process.

```
# error correction model with demeaned spread
spread.ts <- ts(spread-mean(spread))
reg2 <- dynlm( d(stocka.ts) ~ 0 + spread.ts + L(d(stockh.ts))+ L(d(stocka.ts)) )
summary(reg2)
```

Error correction model estimated above:

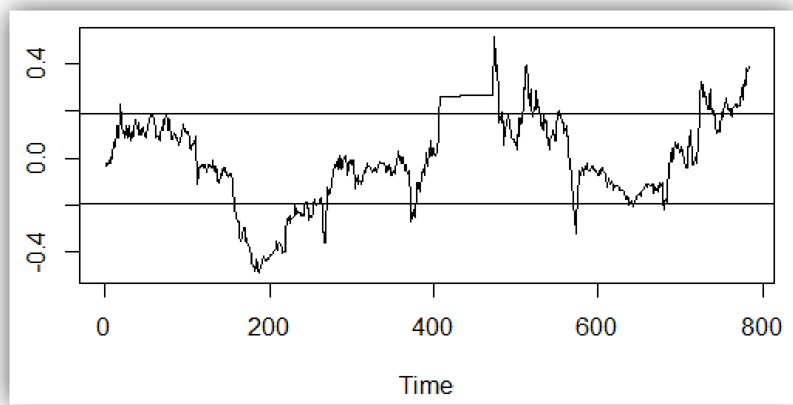
$$\Delta stocka_t = \theta_0(spread_t - spread\ mean) \\ + \theta_1 \Delta stocka_{t-1} + \theta_2 \Delta stockh_{t-1} + u_t$$

- **Step 5:** Plot the stationary spread series with bounds and discuss implications

It's also easy to do plotting in R, we can get visual output of our result by making plots.

```
# demeaned spread series
plot(spread.ts)
abline(sd(spread),0)
abline(-sd(spread),0)
```

Output:



If you were lucky enough, i.e., the spread series are stationary enough, a trading strategy results!

Assuming that the trading cost is less than the gap between spread bounds, a pairs trading strategy comes below:

- Buy a share of A stock and short beta shares of H stock at time t if the spread hits the lower bound
- unwind the position at time later if the spread hits the upper bound