Time series: Cointegration

November 24, 2021

1 Unit Roots and Integration

Univariate time series – unit roots, trends, and stationarity

Have so far glossed over the question of stationarity, except for my stating that for example inflation is stationary whereas the CPI is not stationarity.

We work with two typical processes we can transform before they become stationary.

1. Trend stationary process
$$y_t = \alpha + \beta t + u_t$$

2. Nonstationary model, random walk with drift
$$y_t = \mu + y_{t-1} + u_t$$

The trend-stationary case is easiest to deal with.

If there is a trend, we remove it by a preliminary regression

$$y_t = \alpha + \beta t + e_t$$

and use the residuals e_t of this regression as data. These will be stationary.

In many cases it will be obvious when we have a trend which is to be removed. In other cases it may be hard to distinguish trends and unit roots (See e.g.Campbell and Perron (1991))

We will not say much more about trend removal, instead concentrate on unit roots, since these are of most importance in finance applications.

1.1 Unit roots in the univariate case

Let us first gain some intuition about what unit roots mean?

Exercise 1.

Consider the process
$$w_t = w_{t-1} + \varepsilon_t, \quad w_0 = 0, \quad \varepsilon_t \sim IID(0,1)$$

which is called a *standardized random walk process*. Here ε_t is white noise. Note that this is a recursive process

$$w_t = \sum_{s=1}^{t} \varepsilon_s$$

Show that
$$\mathsf{var}(w_t) = t$$

Is this a covariance stationary process?

Solution to Exercise 1.

$$\mathsf{var}(w_t) = \sum_{s=1}^{t} \mathsf{var}(\varepsilon_t) = \sum_{s=1}^{t} 1 = t \cdot 1 = t$$

where we have used the white noise property of independence after the first equality sign to remove all covariance terms $\mathsf{cov}(\varepsilon_t, \varepsilon_{t+i}) = 0$ if $i \neq 0$. The second step is merely to replace $\mathsf{var}(\varepsilon_t)$ with one, its defined value.

This is not a covariance stationary process.

Exercise 2.

Consider a general AR(1) process without drift.

$$y_t = \phi y_{t-1} + u_t$$

Write y_t as a function of T past errors $u_t, u_{t-1}, u_{t-2}, \cdots u_{t-T}$. What is the effect at time t of a shock at time $t-T$ (i.e. effect of a change in u_{t-T}), for the three cases $\phi < 1, \phi = 1$ and $\phi > 1$? What happens to this effect as $T \to \infty$ (or how long is the *memory* of the process?

Solution to Exercise 2.

Iterate backwards

$$\begin{aligned}
y_t &= \phi y_{t-1} + u_t \\
&= \phi\left(\phi y_{t-2} + u_{t-1}\right) + u_t \\
&= \phi^2 y_{t-2} + \phi u_{t-1} + u_t \\
&= \phi^2\left(\phi y_{t-3} + u_{t-2}\right) + \phi u_{t-1} + u_t \\
&= \phi^3 y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\
&\vdots \\
&= \phi^T y_{t-T} + \phi^{T-1} u_{t-T+1} + \cdots + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\
&= \phi^T\left(\phi y_{t-T-1} + u_{t-T}\right) + \phi^{T-1} u_{t-T+1} + \cdots + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\
&= \phi^{T+1} \phi y_{t-T-1} + \phi^T u_{t-T} + \phi^{T-1} u_{t-T+1} + \cdots + \phi^2 u_{t-2} + \phi u_{t-1} + u_t
\end{aligned}$$

The effect on the observation at time t of a shock at time $t-T$ is then $\phi^T u_{t-T}$.

If $|\phi| < 1$ then $\phi^T \to 0$ as $T - \infty$.
If $\phi = 1$ then $\phi^T = 1$ for all T.
If $\phi > 1$ then $\phi^T \to \infty$ as $T - \infty$.

The case of $\phi > 1$ is explosive, and therefore not a good description of most economic series. We therefore concentrate on the two cases $\phi = 1$ and $|\phi| < 1$. The first is called a unit root.

If we have a unit root the series is not stationary, and can not be used for inference. We need to transform the series somehow. The typical method is to *difference* the series.

1.2 Dealing with unit roots: differencing

Exercise 3.

Suppose we have a unit root proces

$$y_t = y_{t-1} + u_t,$$

where u_t is white noise.

Show that the series of first differences

$$\Delta y_t = y_t - y_{t-1}$$

is covariance stationary

Solution to Exercise 3.

$$y_t = y_{t-1} + u_t,$$
$$y_t - y_{t-1} = (y_{t-1} + u_t) - y_{t-1} = u_t$$

Since u_t is white noise, the series is covariance stationary, since

$$E[u_t] = 0$$

$$\text{var}(u_t) = \sigma^2$$

$$\text{cov}(u_t, u_{t-j}) = 0 \quad \text{if } j \neq 0$$

The important issue in autoregressive processes of this kind is to distinguish

1. Do we have a unt root process that needs to be differenced? or
2. Dow we have a stationary series ($|\phi| < 1$) already, and no differencing is necessary?

Some terminology.

If a nonstationary series must be differenced d times before it becomes stationary, the series is said to be integrated of order d, $I(d)$.

1.2.1 Testing for unit roots

Important issue, testing for whether a process is unit root (I(1)) and therefore must be differenced before it becomes stationary.

Hence, before analysing a time sereis we do pretesting by unit root tests to see whether we need to difference the data.

To test for the presence of a unit root, we can apply either of the tests proposed by Dickey and Fuller.

The original formulation of Dickey-Fuller tests concerns the regression

$$y_t = \beta_0 + \gamma t + \rho y_{t-1} + \varepsilon_t$$

The Dickey - Fuller test is a test of whether $\rho = 1$ in this regression.

The test can alternatively be based on

$$\Delta y_t = \beta_0 + \gamma t + \rho y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \varepsilon_t$$

and is then called the Augmented Dickey Fuller test.

In both cases we can include or not the trend in the tests.

Inference on ρ is made difficult by the fact that under the null, y_t is not stationary.

We have to use critical values based on simulations, Luckily, any software performing Dickey-Fuller tests will typically print out p-values or critical values. Alternatively one can look up a table with critical levels of this test, in most time series books.

Exercise 4.

Consider 1 year risk free interest rates recoverable from treasury securities. Let $R_{1y,t}$ be the 1 year interest rate. Collect estimates of the series for the US. Using data starting in 1962, consider the following issues.

- Fit an AR(1) to the series using OLS.
- Use the Dickey Fuller test statistic to test for a unit root.
- Use the Augmented Dickey Fuller test statistic with 2 lags to test for a unit root.
- Use the Augmented Dickey Fuller test statistic with optimal number of lags to test for a unit root.

Solution to Exercise 4.

Reading the data

```
library(zoo)
library(xts)
TS2treas <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15.csv",
                     skip=6,header=TRUE,na.strings="ND",sep="\t",format="%Y-%m-%d")
dtreas <- as.xts(TS2treas,RECLASS=TRUE)
mtreas <- na.locf(dtreas)[endpoints(na.locf(dtreas),on="months")]
Tbill3m <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15_3MTbill.csv",
                    skip=5,header=TRUE,na.strings="ND",sep=",",format="%Y-%m-%d")
mtbill3m <- na.locf(Tbill3m)[endpoints(na.locf(Tbill3m),on="months")]
mtbill3m <- as.xts(mtbill3m,RECLASS=TRUE)
m1y <- mtreas$Y1
data <- merge(mtbill3m,m1y,all=FALSE)
mtbill3m <- data$mtbill3m
m1y <- data$Y1
```

First doing an estimation of the AR(1)

```
> ar(m1y,order.max=1)

Call:
ar(x = m1y, order.max = 1)

Coefficients:
     1
0.9837

Order selected 1  sigma^2 estimated as  0.3237
```

Then the Dickey Fuller tests

The simple Dickey fuller specifies $k = 0$

```
> library(tseries)
> adf.test(as.matrix(m1y),k=0)

Augmented Dickey-Fuller Test

data:  as.matrix(m1y)
Dickey-Fuller = -2.2929, Lag order = 0, p-value = 0.4543
alternative hypothesis: stationary
```

We do not reject

with two lags

```
> adf.test(as.matrix(m1y),k=2)

Augmented Dickey-Fuller Test

data:  as.matrix(m1y)
Dickey-Fuller = -2.5057, Lag order = 2, p-value = 0.3642
alternative hypothesis: stationary
```

We do not reject

Optimal, by not specifying k

```
> adf.test(as.matrix(m1y))
Augmented Dickey-Fuller Test
data:  as.matrix(m1y)
Dickey-Fuller = -2.2756, Lag order = 8, p-value = 0.4617
alternative hypothesis: stationary
```

We do not reject

A general feature of unit root tests are that they have low power, they may have a hard time distinguishing an unit root and a process that is *close* to a unit root (near integrated, fractionally integrated). Threre are a number of alternative tests of unit roots with better power, but none of them are very good.

There is also an issue that unit root series are almost impossible to distinguish from a stationary series with one or several *breaks* in the time series relations. (structural breaks).

References Textbook: Brooks (2002)

With some examples from Stock and Watson (2019)

2 Summarizing Unit Roots

Trend stationary process

$$y_t = \alpha + \beta t + u_t$$

Nonstationary model, random walk with drift (unit root)

$$y_t = \mu + y_{t-1} + u_t$$

If you have a unit root (integrated of order 1), need to difference once before series is stationary.

If a nonstationary series must be differenced d times before it becomes stationary, the series is said to be integrated of order d, $I(d)$.

Testing for unit roots: Investigating ρ in either of

$$y_t = \beta_0 + \gamma t + \rho y_{t-1} + \varepsilon_t \text{Dickey-Fuller}$$

$$\Delta y_t = \beta_0 + \gamma t + \rho y_{t-1} + \varepsilon_t \text{Augmented Dickey Fuller}$$

Critical values: nonstandard

3 Cointegration

In the univariate case, whenever we find a unit root we will difference the data and work with them in differenced form.

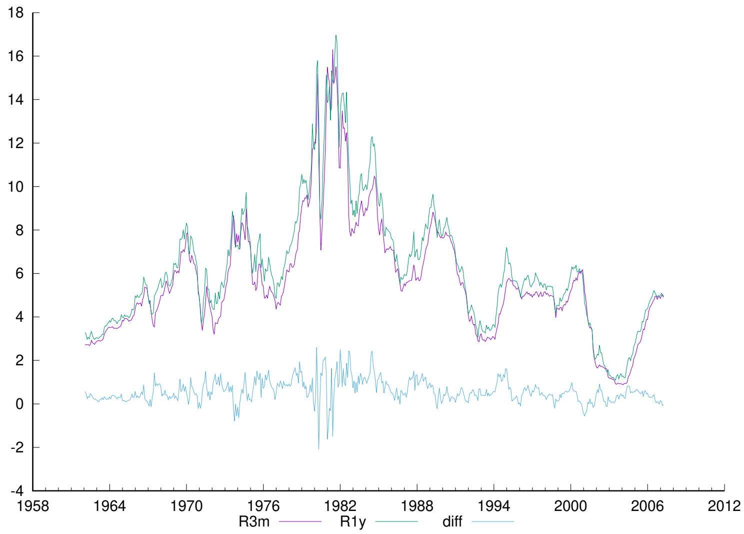
You'd think this will lead you to do the same thing if these time series variables enter into multivariate relationships, and you'd think that you could deal with one time series at a time, difference it until it is stationary, and then work with these series.

While this is the procedure in many cases, there is one exception, where we instead will work with the I(1) series. This exception is called cointegration

This is a case where two unit root series are "shadowing" each others.

Use example from Stock and Watson-interest rates of different maturities.

Consider the following picture, showing the levels of two different interest rates, the 3 and 12 month interest rates, and the differnce between them.



In this picture, the two interest rate series move together in lock-step. They are obvious candidates for being cointegrated.

Definition: Suppose X_t and Y_t are integrated of order one. If, for some coefficient θ, $Y_t - \theta X_t$ are integrated of order zero, then X_t and Y_t are said to be *cointegrated*. The coefficient θ is called the *cointegrating coefficient*.

If X_t and Y_t are cointegrated, they have the same, or common, stochastic trend.

Computing the difference $Y_t - \theta X_t$ eliminates the common stochastic trend.

There are three ways of deciding whether two variables can be viewed as cointegrating

1. Expert knowledge and economic theory.
2. Graph series and look for common stochastic trends.
3. Do statistical tests for cointegration.

Let us look at the term structure example

1) Economic theory

$R_{3m,t}$ is the interest rate on 3 month borrowing

$R_{1y,t}$ is the interest rate on 1 year borrowing By definition

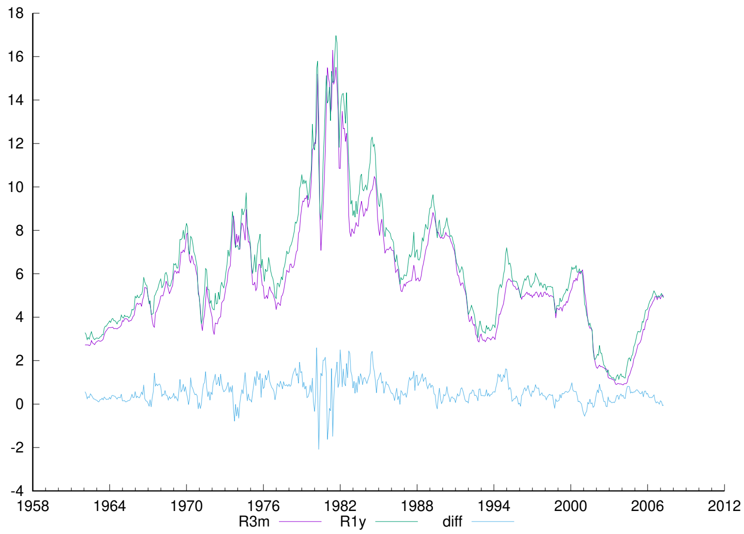
$$(1 + R_{1y,t}) = (1 + R_{3m,t})(1 + f_{3m,1y,t})$$

where $f_{3m,1y,t}$ is the forward rate for borrowing between $3m$ and $1y$.

The expectations theory of interest rates states the forward rate is the excpected 9 month interest rate 3 months from now.

The implication from the expectations theory is that R_{3m} and R_{1y} are cointegrated with cointegrating coefficient of $\theta = 1$.

2) Visual expections



3) Testing for cointegration

a) Testing for cointegration when θ is known.

This is tested by considering the time sereis of $(Y_t - \theta X_t)$ and running standard unit root tests on this time series.

b) Testing for integration when coefficient θ in $Y_t - \theta X_t$ is unknown.

The preocedure is called the Engle-Granger Augmented Dickey-Fuller tests (EG-ADF)

Step 1: Estimate

$$y_t = \alpha + \beta x_t + z_t$$

Apply the Dickey-Fuller procedure to the residuals z_t of this regression.

Exercise 5.

Consider the link between interest rates at different points in the term structure, for example. comparing the 3 month and 1 year risk free interest rates recoverable from treasury securities. Between these two interest rates we expect to find clear relationships, due to the expectations theory of the term structure.

Let $R_{3m,t}$ be the three month interest rate and $R_{1y,t}$ the 1 year interest rate. Collect estimates of these series for the US for as long time periods as is possible. What is the period available? (Hint: the Federal Reserve.)

Consider the following issues.

- Looking at these two series as univariate series, do the interest rates have unit roots?
- Plot the "term spread," the difference between the two interest rate series. Is there reason to believe these two series are cointegrated?
- Suppose you know the cointegrating coefficient $\theta = 1$. Test for cointegration.
- Suppose you don't know the cointegrating coefficient θ in the relationship

$$diff = R_{1y,t} - \theta R_{3m,t}$$

Test for cointegration.

Solution to Exercise 5.

Going to the fed, you find 3 month treasury series starting in 1934 and 1 year treasury series starting in 1962. To investigate the joint behavious of these, it is the 1 year series which is the limit. Use data starting in 1962 for the investigation.

Read the data

```
library(zoo)
library(xts)
TS2treas <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15.csv",
                     skip=6,header=TRUE,na.strings="ND",sep="\t",format="%Y-%m-%d")
dtreas <- as.xts(TS2treas)
mtreas <- na.locf(dtreas)[endpoints(na.locf(dtreas),on="months")]
Tbill3m <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15_3MTbill.csv",
                    skip=5,header=TRUE,na.strings="ND",sep=",",format="%Y-%m-%d")
mtbill3m <- na.locf(Tbill3m)[endpoints(na.locf(Tbill3m),on="months")]
mtbill3m <- as.xts(mtbill3m)
```

Note that I had to do do some massaging of the data to go from daily to end-of-month observations, using the *xts* time series method.

First, just calculate the AR(1) model using OLS on the raw interest series.

First, the 3 month series

```
> ar(mtbill3m,order.max=1)

Coefficients:
     1
0.9832
Order selected 1  sigma^2 estimated as  0.2902
```

Doing the same for the 1 year series.

```
> ar(m1y,order.max=1)
Coefficients:
     1
0.9837
Order selected 1  sigma^2 estimated as  0.3237
```

Both coefficients on the AR(1) term is close to one, and warrants further investigation as to whether there is a unit root in the series.

Running Dickey-Fuller tests, for the 3 month case

```
library(tseries)
> adf.test(as.matrix(mtbill3m),k=0)

Augmented Dickey-Fuller Test

data:  as.matrix(mtbill3m)
Dickey-Fuller = -2.3602, Lag order = 0, p-value = 0.4258
alternative hypothesis: stationary

> adf.test(as.matrix(mtbill3m),k=1)

Augmented Dickey-Fuller Test

data:  as.matrix(mtbill3m)
Dickey-Fuller = -2.6645, Lag order = 1, p-value = 0.297
alternative hypothesis: stationary
```

```
> adf.test(as.matrix(mtbill3m))

Augmented Dickey-Fuller Test

data:  as.matrix(mtbill3m)
Dickey-Fuller = -2.4132, Lag order = 8, p-value = 0.4034
alternative hypothesis: stationary
```

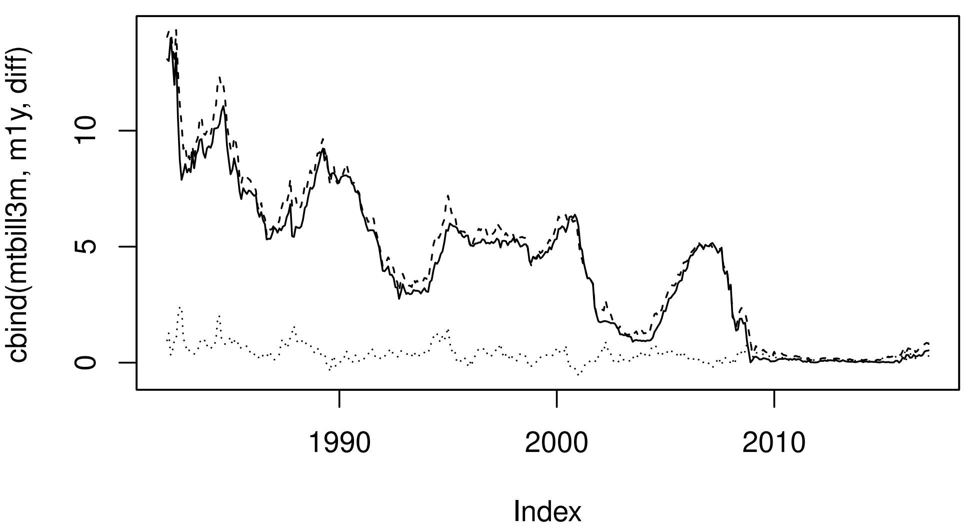
find for the 1 year case

```
> adf.test(as.matrix(m1y),k=0)

Augmented Dickey-Fuller Test

data:  as.matrix(m1y)
Dickey-Fuller = -2.2929, Lag order = 0, p-value = 0.4543
alternative hypothesis: stationary

> adf.test(as.matrix(m1y),k=1)

Augmented Dickey-Fuller Test

data:  as.matrix(m1y)
Dickey-Fuller = -2.7373, Lag order = 1, p-value = 0.2662
alternative hypothesis: stationary

> adf.test(as.matrix(m1y))

Augmented Dickey-Fuller Test

data:  as.matrix(m1y)
Dickey-Fuller = -2.2756, Lag order = 8, p-value = 0.4617
alternative hypothesis: stationary
```

In neither case do we reject the null

We plot the two time series together with the term spread, defined as

$$diff = R_{1y,t} - R_{3m,t}$$


cbind(mtbill3m, m1y, diff)
0
5
10
1990
2000
2010
Index

Testing for I(0) of difference $diff = R_{1y,t} - R_{3m,t}$:

```
> diff <- m1y-mtbill3m
> adf.test(as.matrix(diff),k=0)

Augmented Dickey-Fuller Test

data:  as.matrix(diff)
Dickey-Fuller = -7.6649, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(as.matrix(diff), k = 0) : p-value smaller than printed p-value
> adf.test(as.matrix(diff),k=1)

Augmented Dickey-Fuller Test

data:  as.matrix(diff)
Dickey-Fuller = -6.8379, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(as.matrix(diff), k = 1) : p-value smaller than printed p-value
> adf.test(as.matrix(diff))

Augmented Dickey-Fuller Test

data:  as.matrix(diff)
Dickey-Fuller = -4.5498, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(as.matrix(diff)) : p-value smaller than printed p-value
```

Here we reject the null of unit root, and accept that the two series are cointegrated.

Now, to test for cointegration when estimating the relationship between the cointegrating variables. First run regression

$$R_{1y,t} = \alpha + \beta R_{3m,t} + \varepsilon_t$$

and then do a Dickey-Fuller type of test

First, regression output.

```
> reg <- lm(as.matrix(m1y)~as.matrix(mtbill3m))
> summary(reg)

Call:
lm(formula = as.matrix(m1y) ~ as.matrix(mtbill3m))

Residuals:
     Min       1Q   Median       3Q      Max
-1.75194 -0.23226 -0.04908  0.21498  1.94184

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.230783   0.033985   6.791 2.69e-11 ***
as.matrix(mtbill3m) 1.063822   0.005679 187.315  < 2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4103 on 599 degrees of freedom
Multiple R-squared: 0.9832,Adjusted R-squared: 0.9832
F-statistic: 3.509e+04 on 1 and 599 DF,  p-value: < 2.2e-16
```

While economic theory suggests

$$R_{1y,t} = R_{3m,t}$$

we estimate

$$R_{1y,t} = 0.23 + 1.06 R_{3m,t}$$

And then do Dickey Fuller tests on the residuals.

In the following do various DF variations.

```
> res <- as.matrix(residuals(reg))
> adf.test(res,k=0)

Augmented Dickey-Fuller Test

data:  res
Dickey-Fuller = -8.6055, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(res, k = 0) : p-value smaller than printed p-value
> adf.test(res,k=1)

Augmented Dickey-Fuller Test

data:  res
Dickey-Fuller = -7.6247, Lag order = 1, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(res, k = 1) : p-value smaller than printed p-value
> adf.test(res)

Augmented Dickey-Fuller Test

data:  res
Dickey-Fuller = -5.5397, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(res) : p-value smaller than printed p-value
```

We conclude that we reject the unit root hypothesis for the difference, term spread is I(0) even if the interest rates are both I(1)

3.1 Error correction systems estimation

Given that we have found two variables to be cointegrated, how do we use these relationships in estimation?

The typical next step is to set up an *error correction model*.

If $(Y_t - \theta X_t)$ are cointegrated, in the error correction model this term is added as an explanatory variable.

Exercise 6.

Consider the link between interest rates at different points in the term structure, for example. comparing the 3 month and 1 year risk free interest rates recoverable from treasury securities. Between these two interest rates we expect to find clear relationships, due to the expectations theory of the term structure. Let $R_{3m,t}$ be the three month interest rate and $R_{1y,t}$ the 1 year interest rate. We have seen that using US data for 1962:1 to 2006:12, these two variables are cointegrated.

For predicting these two interest rates, we can add an "error correction term" to estimation, as lagged values of $R_{3m,t} - \theta R_{1y,t}$ are stationary relative to changes in interest rates.

Consider the following system

$$
\begin{aligned}
\Delta R_{3m,t} &= \mu_1 + \beta_{11}\Delta R_{3m,t-1} + \beta_{12}\Delta R_{3m,t-2} \\
&\quad +\gamma_{11}\Delta R_{1y,t-1} + \gamma_{12}\Delta R_{1y,t-2} \\
&\quad +\phi_1(R_{1y,t-1} - \theta R_{3y,t-1}) \\
\Delta R_{1y,t} &= \mu_1 + \gamma_{21}\Delta R_{3m,t-1} + \gamma_{22}\Delta R_{3m,t-2} \\
&\quad +\beta_{21}\Delta R_{1y,t-1} + \beta_{22}\Delta R_{1y,t-2} \\
&\quad +\phi_2(R_{1y,t-1} - \theta R_{3y,t-1})
\end{aligned}
$$

Using economic theory, what is the value of θ?

Using this value for θ, estimate the model using OLS equation by equation. Are the error correction terms important predictors?

Solution to Exercise 6.

Reading the data

```
library(zoo)
library(xts)
TS2treas <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15.csv",
                     skip=6,header=TRUE,na.strings="ND",sep="\t",format="%Y-%m-%d")
dtreas <- as.xts(TS2treas,RECLASS=TRUE)
mtreas <- na.locf(dtreas)[endpoints(na.locf(dtreas),on="months")]
Tbill3m <- read.zoo("../../../../data/usa/us_treasuries/FRB_H15_3MTbill.csv",
                    skip=5,header=TRUE,na.strings="ND",sep=",",format="%Y-%m-%d")
mtbill3m <- na.locf(Tbill3m)[endpoints(na.locf(Tbill3m),on="months")]
mtbill3m <- as.xts(mtbill3m,RECLASS=TRUE)
m1y <- mtreas$Y1
data <- merge(mtbill3m,m1y,all=FALSE)
mtbill3m <- data$mtbill3m
m1y <- data$Y1
```

Use library dynlm

```
> library(dynlm)
```

First estimate the model equation by equation without the error correction term.

The equation with R_{3m} as dependent variable.

```
> reg3m <- dynlm(d(mtbill3m) ~ L(d(mtbill3m),1) + L(d(mtbill3m),2) + L(d(m1y),1) + L(d(m1y),2) )
> summary(reg3m)

Time series regression with "zoo" data:
Start = 1962-04-30, End = 2012-02-29

Call:
dynlm(formula = d(mtbill3m) ~ L(d(mtbill3m), 1) + L(d(mtbill3m),
    2) + L(d(m1y), 1) + L(d(m1y), 2))

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-3.6440 -0.1178  0.0160  0.1651  2.7549

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.003855   0.019500  -0.198   0.8434
L(d(mtbill3m), 1) -0.145584   0.079872  -1.823   0.0689 .
L(d(mtbill3m), 2)  0.067879   0.079562   0.853   0.3939
L(d(m1y), 1)       0.325404   0.075349   4.319 1.84e-05 ***
L(d(m1y), 2)      -0.137359   0.076284  -1.801   0.0723 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4768 on 593 degrees of freedom
Multiple R-squared: 0.05622,Adjusted R-squared: 0.04985
F-statistic: 8.831 on 4 and 593 DF,  p-value: 6.26e-07
```

Regression using one year as dependent variable

```
> reg1y <- dynlm(d(m1y) ~ L(d(mtbill3m),1) + L(d(mtbill3m),2) + L(d(m1y),1) + L(d(m1y),2) )
> summary(reg1y)

Time series regression with "zoo" data:
Start = 1962-04-30, End = 2012-02-29

Call:
dynlm(formula = d(m1y) ~ L(d(mtbill3m), 1) + L(d(mtbill3m), 2) +
    L(d(m1y), 1) + L(d(m1y), 2))

Residuals:
    Min      1Q  Median      3Q     Max
-4.3000 -0.1696  0.0073  0.2024  2.8418

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.004193   0.020539  -0.204  0.83832
L(d(mtbill3m), 1)  0.043345   0.084130   0.515  0.60660
L(d(mtbill3m), 2)  0.178821   0.083803   2.134  0.03327 *
L(d(m1y), 1)       0.168588   0.079365   2.124  0.03407 *
L(d(m1y), 2)      -0.263685   0.080350  -3.282  0.00109 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5022 on 593 degrees of freedom
Multiple R-squared: 0.05345,Adjusted R-squared: 0.04706
F-statistic: 8.371 on 4 and 593 DF,  p-value: 1.424e-06
```

Then redo this with the difference between interest rates as explanatory variable

The equation with R_{3m} as dependent variable.

```
> reg3m <- dynlm(d(mtbill3m) ~ L(d(mtbill3m),1) + L(d(mtbill3m),2) + L(d(m1y),1) + L(d(m1y),2) + L(m1y-mtbill3m) )
> summary(reg3m)

Time series regression with "zoo" data:
```

```
Start = 1962-04-30, End = 2012-02-29

Call:
dynlm(formula = d(mtbill3m) ~ L(d(mtbill3m), 1) + L(d(mtbill3m),
    2) + L(d(m1y), 1) + L(d(m1y), 2) + L(m1y - mtbill3m))

Residuals:
    Min      1Q  Median      3Q     Max
-3.6576 -0.1181  0.0220  0.1639  2.7730

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       -0.02150    0.03292  -0.653 0.514017
L(d(mtbill3m), 1) -0.12715    0.08458  -1.503 0.133268
L(d(mtbill3m), 2)  0.08143    0.08217   0.991 0.322047
L(d(m1y), 1)       0.30797    0.07981   3.859 0.000126 ***
L(d(m1y), 2)      -0.15144    0.07920  -1.912 0.056341 .
L(m1y - mtbill3m)  0.03124    0.04696   0.665 0.506117
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.477 on 592 degrees of freedom
Multiple R-squared: 0.05692,Adjusted R-squared: 0.04896
F-statistic: 7.147 on 5 and 592 DF,  p-value: 1.662e-06
```

The equation with R_{1y} as dependent variable.

```
> reg1y <- dynlm(d(m1y) ~ L(d(mtbill3m),1) + L(d(mtbill3m),2) + L(d(m1y),1) + L(d(m1y),2) + L(m1y-mtbill3m) )
> summary(reg1y)

Time series regression with "zoo" data:
Start = 1962-04-30, End = 2012-02-29

Call:
dynlm(formula = d(m1y) ~ L(d(mtbill3m), 1) + L(d(mtbill3m), 2) +
    L(d(m1y), 1) + L(d(m1y), 2) + L(m1y - mtbill3m))

Residuals:
    Min      1Q  Median      3Q     Max
-4.2481 -0.1710 -0.0151  0.2067  2.7723

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.06350    0.03452   1.840  0.06631 .
L(d(mtbill3m), 1) -0.02738    0.08867  -0.309  0.75765
L(d(mtbill3m), 2)  0.12682    0.08615   1.472  0.14153
L(d(m1y), 1)       0.23549    0.08368   2.814  0.00505 **
L(d(m1y), 2)      -0.20964    0.08304  -2.525  0.01184 *
L(m1y - mtbill3m) -0.11987    0.04923  -2.435  0.01520 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5001 on 592 degrees of freedom
Multiple R-squared: 0.06283,Adjusted R-squared: 0.05492
F-statistic: 7.938 on 5 and 592 DF,  p-value: 2.977e-07
```

>

Note that in the equation for the one year rate the coefficient on the error correction term is significant.

References Textbook: Brooks (2002)
With some examples from Stock and Watson (2019)

4 Cointegration-summary

Y_t and X_t are I(1) processes. $(Y_t - \theta X_t)$ an I(0) process. Then X_t and Y_t are *cointegrated*
Methods for discovering cointegration

1. Expert knowledge and economic theory.
2. Graph series and look for common stochastic trends.
3. Do statistical tests for cointegration. Engle-Granger Augmented Dickey Fuller - EG-ADF.

Error correction systems modelling: If X_t and Y_t are cointegrated, using lagged values of $(Y_t - \theta X_t)$ as explanatory variables.

References

Chris Brooks. *Introductory econometrics for finance.* Cambridge, 2002.

John Y Campbell and Pierre Perron. Pitfalls and opportunities: What Macroeconmists should know about Unit Roots. In Olivier Jean Blanchard and Stanley Fisher, editors, *NBER Macroeconomics annual 1991.* MIT Press, 1991.

James H Stock and Mark W Watson. *Introduction to Econometrics.* Pearson, 4th edition, 2019.