Chapter 7

Cointegration

7.1 Introduction

Let $\{X_t : t \in \mathbb{Z}\}$ be a time series, If $\{\nabla^d X_t\}$ is a stationary time series, but $\{\nabla^{d-1} X_t\}$ is not, then $\{X_t\}$ is said to be *integrated of order d*, written more concisely as $\{X_t\} \sim I(d)$.

Recall that ∇ denotes the *difference operator* $\nabla X_t = X_t - X_{t-1}$ and hence that $\nabla = (1-B)$ where B denotes the backward shift operator. The power denotes:

$$\nabla^d := (1-B)^d.$$

If $\{\underline{X}_t\}$ is an m-variate time series, the series $\{\nabla^d \underline{X}_t\}$ is defined as the m-variate time series with jth component $\{\nabla^d X_{tj} : t \in \mathbb{Z}\}$.

An m-variate time series $\{\underline{X}_t : t \in \mathbb{Z}\}$ is integrated of order d if each component of $\nabla^d \underline{X}_t$ is stationary, but each component of $\nabla^{d-1}\underline{X}_t$ is not. This is written: $\{\underline{X}_t\} \sim I(d)$. The $I(d)$ process $\{\underline{X}_t\}$ is said to be *cointegrated with cointegration vector* $\underline{\alpha}$ if $\underline{\alpha}$ is an m-vector such that the univariate time series $\{\underline{\alpha}^t \underline{X}_t\}$ is integrated of order *less than d*.

Example 7.1.

Consider the following bi-variate process:

$$\begin{cases} X_t = \sum_{j=1}^{t} Z_j & t \in 1,2,\dots \quad \{Z_t\} \sim \text{IID}(0,\sigma^2) \\ Y_t = X_t + W_t & t \in 1,2,\dots \quad \{W_t\} \sim \text{IID}(0,\tau^2) \end{cases}$$

where $\{W_t\} \perp \{Z_t\}$. Then clearly $\{(X_t, Y_t)^t\}$ is integrated of order 1, since

$$\nabla X_t = Z_t \qquad \nabla Y_t = Z_t + (W_t - W_{t-1}).$$

It is *cointegrated with cointegration vector* $\underline{\alpha} = (1,-1)^t$, since

$$X_t - Y_t = -W_t$$

which is stationary.

Let $U_t = \nabla X_t = Z_t$ and $V_t = \nabla Y_t = Z_t + W_t - W_{t-1}$. Then the series $(U_t, V_t)^t$ may be expressed as a bivariate MA(1) process:

$$\begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Z_t \\ Z_t + W_t \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} Z_{t-1} \\ Z_{t-1} + W_{t-1} \end{pmatrix}$$

and

$$\begin{pmatrix} Z_t \\ Z_t + W_t \end{pmatrix} \sim \text{WN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \tau^2 \end{pmatrix} \right)$$

The process is not invertible, since

$$\det \begin{pmatrix} 1 & 0 \\ z & 1 - z \end{pmatrix} = 1 - z$$

which is 0 for $z = 1$. □

7.2 Error Correction

A system is cointegrated if there are more unit-root nonstationary components than the number of unit roots. It follows that taking the system of equations obtained by differencing each individual component sufficiently often to achieve stationarity results in the problem of unit roots in the MA matrix polynomial, similar to those encountered by over differencing, which causes difficulties in parameter estimation. A VARMA time series is not invertible if the MA matrix polynomial contains unit roots.

Engle and Granger (1987) discuss an error correction representation for a cointegrated system that overcomes the difficulty of estimating noninvertible VARMA models. The following, which develops the example above, gives all the principles.

Example 7.2.

Continuing with the example above, the series $\{(X_t, Y_t)^t : t \in \mathbb{Z}\}$ can be represented as a VAR(1) process

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} = \begin{pmatrix} Z_t \\ Z_t + W_t \end{pmatrix} = \begin{pmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix}$$

where

$$\begin{pmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix} = \begin{pmatrix} Z_t \\ Z_t + W_t \end{pmatrix} \sim \text{IID} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \tau^2 \end{pmatrix} \right).$$

From this, it follows directly, by subtracting $(X_{t-1}, Y_{t-1})^t$ from both sides, that

$$\begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix}.$$

This is stationary, because $(1,-1)\begin{pmatrix} X_t \\ Y_t \end{pmatrix}$ is stationary and the innovation term is also stationary. Furthermore, the 'MA' part does not have the problem of unit roots, since it is the same as the 'MA' part of the original.

When differencing a VARMA, the aim is therefore to express the differenced equation in such a way that the VMA part remains the same and the terms in the VAR part are modified so that each term is stationary. The rearrangement which does this is given by the equations in the following lemma.

Lemma 7.1. *Let*

$$\underline{X}_t - \sum_{j=1}^{p} \Phi_j \underline{X}_{t-j} = \underline{\mu}_t + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j} \qquad \{\underline{Z}_t\} \sim WN(\underline{0}, \Sigma).$$

where $\underline{\mu}_t$ is a deterministic trend function. Then $\nabla \underline{X}_t$ has representation:

$$\nabla \underline{X}_t = \underline{\mu}_t + \Pi \underline{X}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \nabla \underline{X}_{t-i} + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j} \tag{7.1}$$

where

$$\begin{cases} \Phi_j^* = -\sum_{i=j+1}^{p} \Phi_i & j = 1, \ldots, p-1 \\ \Pi = \Phi_p + \Phi_{p-1} + \ldots + \Phi_1 - I = -\Phi(1). \end{cases} \tag{7.2}$$

Proof

$$\begin{aligned} \nabla \underline{X}_t &= \underline{\mu}_t - \underline{X}_{t-1} + \sum_{j=1}^{p} \Phi_j X_{t-j} + \underline{Z}_t + \sum_{j=1}^{q} \Theta \underline{Z}_{t-j} \\ &= \underline{\mu}_t - (I - \sum_{j=1}^{p} \Phi_j) \underline{X}_{t-1} + \sum_{j=1}^{p} \Phi_j (X_{t-j} - X_{t-1}) + \underline{Z}_t + \sum_{j=1}^{q} \Theta \underline{Z}_{t-j} \\ &= \underline{\mu}_t - (I - \sum_{j=1}^{p} \Phi_j) \underline{X}_{t-1} - \sum_{j=2}^{p} \Phi_j \sum_{k=2}^{j} (X_{t-k+1} - X_{t-k}) + \underline{Z}_t + \sum_{j=1}^{q} \Theta \underline{Z}_{t-j} \\ &= \underline{\mu}_t + \Pi \underline{X}_{t-1} + \sum_{j=1}^{p-1} \Phi_i^* \nabla \underline{X}_{t-i} + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j} \end{aligned}$$

where $\Phi_j^* : j = 1, \ldots, p-1$ and Π are given by Equation (7.2). □

Note that the AR matrices $(\Phi_i)_{i=1}^{p}$ can be recovered from the representation of Equation (7.1) via:

$$\begin{cases} \Phi_1 = I + \Pi + \Phi_1^* \\ \Phi_i = \Phi_i^* - \Phi_{i-1}^* \qquad i = 2, \ldots, p, \qquad (\text{using } \Phi_p^* \equiv 0). \end{cases}$$

If the process $\{\nabla \underline{X}_t\}$ in Equation (7.1) it follows that $\Pi \underline{X}_{t-1}$ is stationary. If Π has rank $k < m$, then it may be decomposed into $\Pi = AB^t$ where A and B are $m \times k$ matrices of full rank. The term $\Pi \underline{X}_{t-1}$

is known as the *error correction term*, which plays a key role in cointegration. An error correction representation is:

$$\nabla \underline{X}_t = \underline{\mu}_t + AB^t \underline{X}_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* \nabla X_{t-i} + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j},$$

where A and B are such $m \times k$ full rank matrices. In the decomposition $\Pi = AB^t$, the matrices A and B are not necessarily unique. For any $k \times k$ orthonormal matrix P,

$$AB^t = APP^tB^t = (AP)(BP)^t = A_*B_*^t.$$

where both A_* and B_* are of rank k.

To choose an appropriate matrix B, a common requirement is:

$$B^t = (I_k | B_1^t)$$

where I_k is the $k \times k$ identity matrix and B_1 is a $(m-k) \times k$ matrix. This may require the elements of $\underline{X}_t$ to be re-ordered so that the first k components all have a unit root.

There are three cases of interest in considering the ECM:

1. Rank$(\Pi) = 0$. This implies $\Pi = 0$ and hence that $\underline{X}_t$ is not cointegrated. The ECM reduces to:

$$\nabla \underline{X}_t = \underline{\mu}_t + \Phi_1^* \nabla \underline{X}_{t-1} + \ldots + \Phi_{p-1}^* \nabla \underline{X}_{t-p+1} + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j}$$

   so that $\nabla \underline{X}_t$ follows a VARMA(p-1,q) model with deterministic trend $\underline{\mu}_t$.

2. Rank$(\Pi) = m$ This implies that $|\Phi(1)| \neq 0$ and $\underline{X}_t$ contains no unit roots; that is $\underline{X}_t$ is $I(0)$. The ECM model is not informative and $\underline{X}_t$ should be studied directly.

3. $0 <$ Rank$(\Pi) = k < m$. In this case, Π can be written as:

$$\Pi = AB^t$$

   where A and B are $m \times k$ matrices with Rank$(A) =$ Rank$(B) = k$. The ECM is:

$$\nabla \underline{X}_t = \underline{\mu}_t + AB^t \underline{X}_{t-1} + \Phi_1^* \nabla \underline{X}_{t-1} + \ldots + \Phi_{p-1}^* \nabla \underline{X}_{t-p+1} + \underline{Z}_t + \sum_{j=1}^{q} \Theta_j \underline{Z}_{t-j}.$$

   This means that $\underline{X}_t$ is cointegrated with k linearly independent cointegrating vectors and $m-k$ unit roots, which correspond to $m-k$ common stochastic trends of $\underline{X}_t$.

The discussion assumes that all unit roots are of multiplicity 1, but it is straightforward to extend it to situations where the unit roots have different multiplicities. $I(0)$ or $I(1)$ processes are the most common in applications.

If the innovations are from a known family of distributions and the number of cointegrating factors k is given, then the error correction model can be estimated by maximum likelihood methods.

If $\underline{X}_t$ is cointegrated with $\text{Rank}(\Pi) = k$, then a simple way to obtain the representation of the $m-k$ common trends is to obtain an orthogonal complement matrix $A_\perp$ of A (that, $A_\perp$ is $m \times (m-k)$ such that $A_\perp^t A = 0$) and then set $\underline{Y}_t = A_\perp^t \underline{X}_t$. This is seen by pre-multiplying the ECM by $A_\perp^t$ and using $\Pi = AB^t$.

If $\underline{X}_t$ is at most $I(1)$, it follows that $\nabla \underline{X}_t$ is $I(0)$. If $\underline{X}_t$ contains unit roots, then $\det(\Phi(1)) = 0$ and $\Pi = -\Phi(1)$ is singular.

There is no error correction term in the resulting equation. It follows that the $m-k$-variate process $\underline{Y}_t$ has $m-k$ unit roots.

Example 7.3.

Let $\underline{X}_t$ be the bivariate ARMA(1,1) process defined by:

$$\begin{pmatrix} X_{t1} \\ X_{t2} \end{pmatrix} - \begin{pmatrix} 0.5 & -1.0 \\ -0.25 & 0.5 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} = \begin{pmatrix} Z_{t1} \\ Z_{t2} \end{pmatrix} + \begin{pmatrix} -0.2 & 0.4 \\ 0.1 & -0.2 \end{pmatrix} \begin{pmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{pmatrix}.$$

$$\{\underline{Z}_t\} \sim \text{WN}(\underline{0}, \Sigma) \qquad \det(\Sigma) > 0.$$

This is not a stationary model;

$$\Phi(z) = \begin{pmatrix} 1 - 0.5z & z \\ 0.25z & 1 - 0.5z \end{pmatrix}$$

$$|\Phi(z)| = (1 - 0.5z)^2 - 0.25z^2 = 1 - z$$

which has a unit root. This may be re-written as:

$$\begin{aligned} \begin{pmatrix} \nabla X_{t1} \\ \nabla X_{t2} \end{pmatrix} &= \begin{pmatrix} -0.5 & -1.0 \\ -0.25 & -0.5 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} Z_{t1} \\ Z_{t2} \end{pmatrix} + \begin{pmatrix} -0.2 & 0.4 \\ 0.1 & -0.2 \end{pmatrix} \begin{pmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ -0.5 \end{pmatrix} \begin{pmatrix} 0.5 & 1.0 \end{pmatrix} \begin{pmatrix} X_{t-1,1} \\ X_{t-1,2} \end{pmatrix} + \begin{pmatrix} Z_{t1} \\ Z_{t2} \end{pmatrix} + \begin{pmatrix} -0.2 & 0.4 \\ 0.1 & -0.2 \end{pmatrix} \begin{pmatrix} Z_{t-1,1} \\ Z_{t-1,2} \end{pmatrix} \end{aligned}$$

This is a stationary model because both $\nabla \underline{X}_t$ and $(0.5, 1.0)\underline{X}_t$ are stationary. The MA matrix polynomial does not have unit roots. In this case,

$$A = \begin{pmatrix} -1 \\ -0.5 \end{pmatrix}, \qquad A_\perp = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

giving

$$Y_t = \begin{pmatrix} 1 & -2 \end{pmatrix} \underline{X}_t = X_{t1} - 2X_{t2}$$

□

7.3 Cointegration Test

Consider an m-dimensional VAR(p) time series $\{\underline{X}_t\}$ with possible time trend, so that the model is:

$$\underline{X}_t = \underline{\mu}_t + \Phi_1 \underline{X}_{t-1} + \ldots + \Phi_p \underline{X}_{t-p} + \underline{Z}_t \qquad \{\underline{Z}_t\} \sim N(\underline{0}, \Sigma)$$

The innovations are assumed to be Gaussian. Suppose that

$$\underline{\mu}_t = \underline{\mu}_0 + t\underline{\mu}_1$$

where $\underline{\mu}_0$ and $\underline{\mu}_1$ are m-vectors (the trend is linear). For a specified deterministic term $\underline{\mu}_t$, the rank of Π may be tested using a *maximum likelihood test*. Let $H(k)$ denote the *null* hypothesis that the rank of the matrix Π is k. For example, under $H(0)$, $\text{Rank}(\Pi) = 0$ so that $\Pi = 0$ and there is no cointegration. The hypotheses of interest are:

$$H(0) \subset \ldots \subset H(k) \subset \ldots \subset H(m).$$

For testing purposes, the ECM is written:

$$\nabla \underline{X}_t = \underline{\mu}_0 + t\underline{\mu}_1 + \Pi \underline{X}_{t-1} + \Phi_1^* \nabla \underline{X}_{t-1} + \ldots + \Phi_{p-1}^* \nabla \underline{X}_{t-p+1} + \underline{Z}_t \qquad t = p+1, \ldots T$$

The rank of Π is the number of non zero eigenvalues of Π, which can be obtained if a consistent estimate of Π is available. Clearly Π is related to the *covariance* matrix between $\underline{X}_{t-1}$ and $\nabla \underline{X}_t$ after adjusting for the effects of $\underline{\mu}_t = \underline{\mu}_0 + t\underline{\mu}_1$ and $\nabla \underline{X}_{t-i}, \quad i = 1, \ldots, p-1$. Let

$$\widehat{\underline{U}}_t = \nabla \underline{X}_t - P(\nabla \underline{X}_t | \mathcal{F}(\nabla \underline{X}_{t-i} : i = 1, \ldots, p, \underline{\mu}_0 + t\underline{\mu}_1))$$

and

$$\widehat{\underline{V}}_t = \underline{X}_{t-1} - P(\underline{X}_{t-1} | \mathcal{F}(\nabla \underline{X}_{t-i} : i = 1, \ldots, p, \underline{\mu}_0 + t\underline{\mu}_1))$$

then

$$\widehat{\underline{U}}_t = \Pi \widehat{\underline{V}}_t + \underline{Z}_t.$$

which is the equation of interest for the cointegration test.

Under the normality assumption, the likelihood ratio test for testing the rank of Π can be done using *canonical correlation analysis* between $\widehat{\underline{U}}_t$ and $\widehat{\underline{V}}_t$. These canonical correlations are the *partial canonical* *correlations* between $\nabla \underline{X}_{t-1}$ and $\underline{X}_{t-1}$. Let $\{\widehat{\lambda}_i\}$ denote the squared canonical correlations between $\underline{\widehat{U}}_t$ and $\underline{\widehat{V}}_t$.

Consider the hypotheses:

$$H_0 : \text{Rank}(\Pi) = k \qquad \text{versus} \qquad H_1 : \text{Rank}(\Pi) > k$$

where H_1 denotes the alternative. The *likelihood ratio test*, proposed by Johansen (1988), is:

$$\text{LR}_{\text{tr}}(k) = -(n-p) \sum_{i=k+1}^{m} \ln(1 - \widehat{\lambda}_i)$$

If $\text{Rank}(\Pi) = k$, then $\widehat{\lambda}_i$ should be small for $i > k$ and hence the test statistic should be small. This test is referred to as the *trace* cointegration test.

The distribution of the test statistic under the null hypothesis is not asymptotically chi-squared; the situation does not admit splitting into i.i.d. variables and it is not possible to appeal to a central limit theorem effect. Information on the distribution is obtained via simulation and these values are stored in the routines used to test the hypotheses.

Johansen (1988) also considers a *sequential* procedure to determine the number of cointegrating vectors. Specifically, the hypotheses of interest are:

$$H_0 : \text{Rank}(\Pi) = k \qquad \text{versus} \qquad H_1 : \text{Rank}(\Pi) = k+1.$$

The LR ratio test statistic in this case, called the *maximum eigenvalue statistic*, is:

$$\text{LR}_{\text{max}}(k) = -(n-p) \ln\left(1 - \widehat{\lambda}_{k+1}\right).$$

Again, the statistic does not have a standard distribution and the critical values have been obtained by simulation and are stored in the routines used for testing.

7.4 Forecasting of Conintegrated VAR Models

The fitted ECM can be used to produce forecasts. First, conditioned on the estimated parameters, the ECM equation can be used to produce forecasts of the differenced series $\nabla \underline{X}_t$. These forecasts are then used in turn to obtain forecasts of $\underline{X}_t$. The difference between these forecasts and the VAR forecasts considered earlier is that the ECM approach imposes the cointegration relationships in producing the forecasts.

Example The data for two weekly U.S. short-term interest rates are from December 12, 1958 and August 6, 2004 are found in `w-tb3n6ms.txt`. They are the 3 month and 6 month treasury bill (TB) rates. Firstly, when the Augmented Dickey-Fuller test is applied, the null hypothesis of 'no unit roots' is not rejected for either the series. The 3 month gives a p value of 0.2573, while the 6 month gives a p value of 0.2907. We may therefore proceed on the assumption that these series have unit roots. Secondly, from the plots found in Figure 7.1 show remarkable similarities.

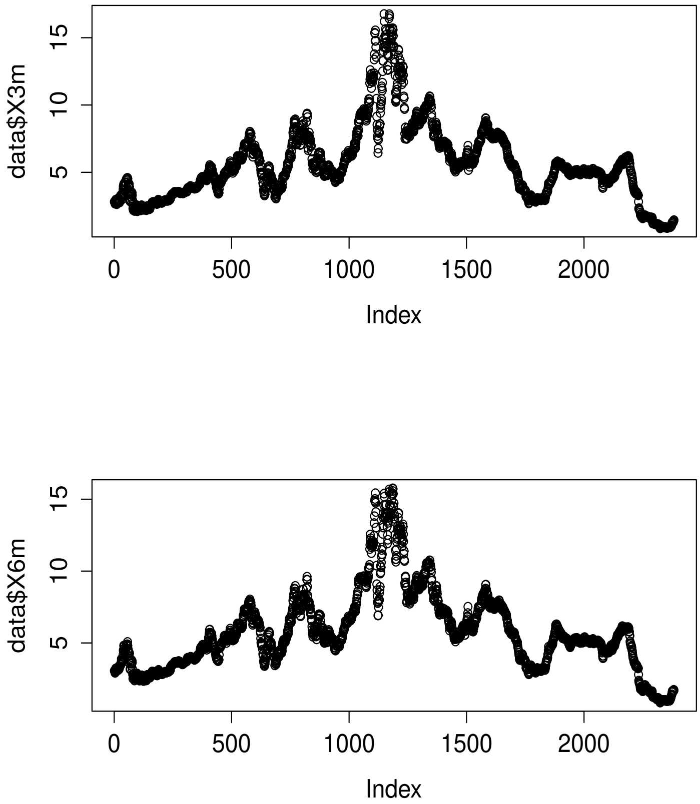

Figure 7.1: Treasury bill 3 month and 6 month

At a rough guess, the difference between the 3 month and 6 month may produce a series that does not have unit roots. The difference (3 month - 6 month) is plotted in Figure 7.2
Applying the Augmented Dickey-Fuller test to the differences gives a Dickey-Fuller statistic of -8.6174 and a p-value of less than 0.01, so the presence of a unit root is *rejected* at the 0.01 significance level; it is safe to work under the assumption that this series of differences does not have a unit root.

Firstly, note that for the co-integration test, the AR order p is *given*. The first thing to do is to find a suitable order. The package `vars` helps with this and should be installed and activated.

```
library(vars)
www =
"https://www.mimuw.edu.pl/~noble/courses/TimeSeries/data/w-tb3n6ms.
txt"
```

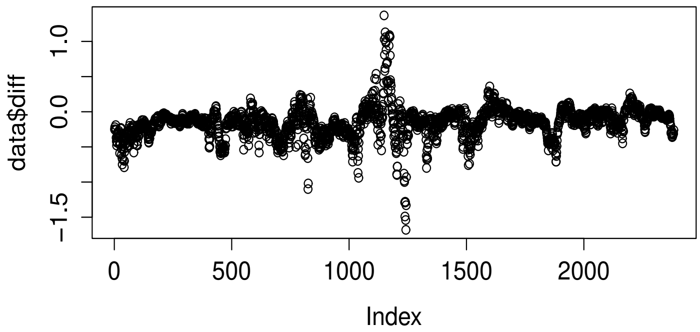

Figure 7.2: Difference: treasury bill 3 month minus 6 month

```
wtb3n6ms = read.table(www,header=T)
data <- wtb3n6ms
x <- cbind(data$X3m, data$X6m)
y <- data.frame(x)
z<-VAR(y,p=4,type="both",ic="AIC")
```

The default value for p is 1, so it is necessary to write in the highest order that one is prepared to accept. The 'both' means both $\underline{\mu}_0$ and $\underline{\mu}_1$ should be estimated for a trend $\underline{\mu}_0 + t\underline{\mu}_1$.

```
> summary(z)
```

The significance levels of the parameter estimates indicate that a VAR(3) model should be appropriate, from the 'estimation results for equation X1'. The constant and the trend are not significant. A VAR(2) model may be OK.

Now perform a cointegration test. A useful package for this is `urca`, which should be installed and activated. This package enables a Johansen test to be performed:

```
?ca.jo
```

gives the syntax for the command `ca.jo`.

```
> cointeigen <- ca.jo(y,type = "eigen", ecdet = "none", K=3)
```

$K = 3$ denotes that the test is based on a VAR(3) model, 'none' indicates that $\underline{\mu}_0 = 0$ and $\underline{\mu}_1 = 0$, and 'eigen' denotes the eigenvalue test described above.

```
> summary(cointeigen)
```

For the trace test,

```
> cointtrace <- ca.jo(y,type="trace",ecdet="none",K=3)
> summary(cointtrace)
```

Both Johansen's tests confirm that the two series are cointegrated with one cointegrating vector when a VAR(3) model is used.

One way to obtain the model is to try the package `tsDyn`

```
> library("tsDyn")
> answer <- VECM(y,2,r=1,include="none")
> summary(answer)
```

The output gives the model:

$$\begin{aligned} \begin{pmatrix} \nabla X_{t1} \\ \nabla X_{t2} \end{pmatrix} &= \begin{pmatrix} -0.091 \\ -0.031 \end{pmatrix} (X_{t1} - 0.98 X_{t2}) \\ &\quad + \begin{pmatrix} 0.045 & 0.268 \\ -0.037 & 0.311 \end{pmatrix} \begin{pmatrix} \nabla X_{t-1,1} \\ \nabla X_{t-1,2} \end{pmatrix} + \begin{pmatrix} -0.208 & 0.258 \\ -0.029 & 0.094 \end{pmatrix} \begin{pmatrix} \nabla X_{t-2,1} \\ \nabla X_{t-2,2} \end{pmatrix} \end{aligned}$$

An output with lag 3 would show that the lag 3 coefficients were not significant.

Finally, we use the fitted ECM to produce 1-step and 10-step-ahead forecasts for both $\nabla \underline{X}_t$ and $\underline{X}_t$. This may be done in the following way:

```
> model <- vec2var(cointeigen)
> pred <- predict(model, n.ahead = 10, ci = 0.95)
> pred$fcst
```

These are the forecasts for the original data.