15

Cointegration

15.1 Introduction

Cointegration analysis is a technique that is frequently applied in econometrics. In finance it can be used to find trading strategies based on mean-reversion.

Suppose one could find a stock whose price (or log-price) series was stationary and therefore mean-reverting. This would be a wonderful investment opportunity. Whenever the price was below the mean, one could buy the stock and realize a profit when the price returned to the mean. Similarly, one could realize profits by selling short whenever the price was above the mean. Alas, returns are stationary but not prices. We have seen that log-prices are integrated. However, not all is lost. Sometimes one can find two or more assets with prices so closely connected that a linear combination of their prices is stationary. Then, a portfolio using as portfolio weights the *cointegrating vector*, which is the vector of coefficients of this linear combination, will have a stationary price. Cointegration analysis is a means for finding cointegration vectors.

Two time series, $Y_{1,t}$ and $Y_{2,t}$, are cointegrated if each is $I(1)$ but if there exists a λ such that $Y_{1,t} - \lambda Y_{2,t}$ is stationary. For example, the common trends model is that

$$\begin{aligned} Y_{1,t} &= \beta_1 W_t + \epsilon_{1,t}, \\ Y_{2,t} &= \beta_2 W_t + \epsilon_{2,t}, \end{aligned}$$

where β_1 and β_2 are nonzero, the trend W_t common to both series is $I(1)$, and the noise processes $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are $I(0)$. Because of the common trend, $Y_{1,t}$ and $Y_{2,t}$ are nonstationary but there is a linear combination of these two series that is free of the trend so they are cointegrated. To see this, note that if $\lambda = \beta_1/\beta_2$, then

$$\beta_2(Y_{1,t} - \lambda Y_{2,t}) = \beta_2 Y_{1,t} - \beta_1 Y_{2,t} = \beta_2 \epsilon_{1,t} - \beta_1 \epsilon_{2,t} \tag{15.1}$$

is free of the trend W_t and therefore is $I(0)$.

D. Ruppert, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics, 413
DOI 10.1007/978-1-4419-7787-8_15, © Springer Science+Business Media, LLC 2011

The definition of cointegration extends to more than two time series. A d-dimensional multivariate time series is cointegrated of order r if the component series are $I(1)$ but r independent linear combinations of the components are $I(0)$ for some r, $0 < r \leq d$. Somewhat different definitions of cointegration exist, but this one is best for our purposes.

In Section 13.2.4 we saw the danger of spurious regression when the residuals are integrated. This problem should make one cautious about regression with nonstationary time series. However, if Y_t is regressed on X_t and the two series are cointegrated, then the residuals will be $I(0)$ so that least-squares estimator will be consistent.

The Phillips–Ouliaris cointegration test regresses one integrated series on others and applies the Phillips–Perron unit root test to the residuals. The null hypothesis is that the residuals are unit root nonstationary, which implies that the series are *not* cointegrated. Therefore, a small p-value implies that the series *are* cointegrated and therefore suitable for regression analysis. The residuals will still be correlated and so they should be modeled as such; see Section 14.1.

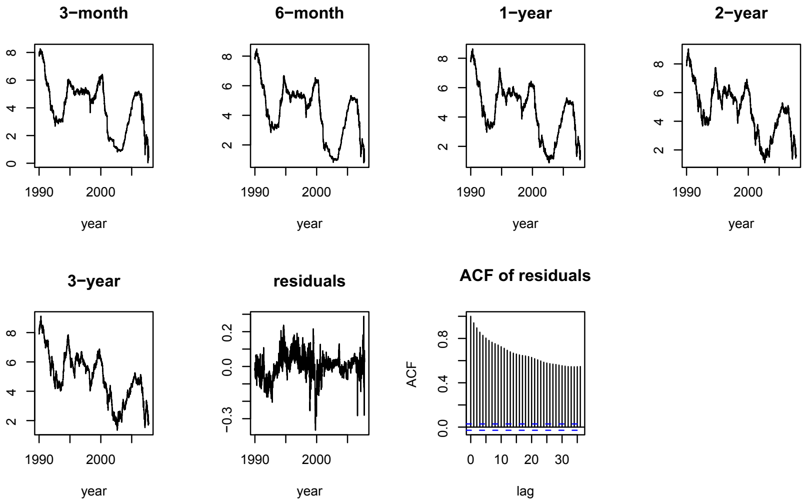

Fig. 15.1. *Time series plots of the five yields and the residuals from a regression of the 1-year yields on the other four yields. Also, a ACF plot of the residuals.*

Example 15.1. Phillips–Ouliaris test on bond yields

This example uses three-month, six-month, one-year, two-year, and three-year bond yields recorded daily from January 2, 1990 to October 31, 2008, for a total of 4714 observations. The five yields series are plotted in Figure 15.1, and one can see that they track each other somewhat closely. This suggests that the five series may be cointegrated. The one-year yields were regressed on the four others and the residuals and their ACF are also plotted in Figure 15.1. The two residual plots are ambiguous about whether the residuals are stationary, so a test of cointegration would be helpful.

Next, the Phillips–Ouliaris test was run using the R function po.test in the tseries package.

```
        Phillips-Ouliaris Cointegration Test

data:  dat[, c(3, 1, 2, 4, 5)]
Phillips-Ouliaris demeaned = -323.546, Truncation lag
parameter = 47, p-value = 0.01

Warning message:
In po.test(dat[, c(3, 1, 2, 4, 5)]) : p-value smaller
than printed p-value
```

The p-value is computed by interpolation if it is within the range of a table in Phillips and Ouliaris (1990). In this example, the p-value is outside the range and we know only that it is below 0.01, the lower limit of the table. The small p-value leads to the conclusion that the residuals are stationary and so the five series are cointegrated.

Though stationary, the residuals have a large amount of autocorrelation and may have long-term memory. They take a long time to revert to their mean of zero. Devising a profitable trading strategy from these yields seems problematic.

□

15.2 Vector Error Correction Models

The regression approach to cointegration is somewhat unsatisfactory, since one series must be chosen as the dependent variable, and this choice must be somewhat arbitrary. In Example 15.1, the middle yield, ordered by maturity, was used but for no compelling reason. Moreover, regression will find only one cointegration vector, but there could be more than one.

An alternative approach to cointegration that treats the series symmetrically uses a *vector error correction model* (VECM). In these models, the deviation from the mean is called the "error" and whenever the stationary linear combination deviates from its mean, then it is pushed back toward its mean (the error is "corrected").

The idea behind error correction is simplest when there are only two series, $Y_{1,t}$ and $Y_{2,t}$. In this case, the error correction model is

$$\Delta Y_{1,t} = \phi_1(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{1,t}, \tag{15.2}$$

$$\Delta Y_{2,t} = \phi_2(Y_{1,t-1} - \lambda Y_{2,t-1}) + \epsilon_{2,t}, \tag{15.3}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are white noises. Subtracting λ times (15.3) from (15.2) gives

$$\Delta(Y_{1,t} - \lambda Y_{2,t}) = (\phi_1 - \lambda\phi_2)(Y_{1,t-1} - \lambda Y_{2,t-1}) + (\epsilon_{1,t} - \lambda\epsilon_{2,t}). \tag{15.4}$$

Let $\mathcal{F}_t$ denote the information set at time t. If $(\phi_1 - \lambda\phi_2) < 0$, then $E\{\Delta(Y_{1,t} - \lambda Y_{2,t})|\mathcal{F}_t\}$ is opposite in sign to $Y_{1,t-1} - \lambda Y_{2,t-1}$. This causes error correction because whenever $Y_{1,t-1} - \lambda Y_{2,t-1}$ is positive, its expected change is negative and vice versa.

A rearrangement of (15.4) shows that $Y_{1,t-1} - \lambda Y_{2,t-1}$ is an AR(1) process with coefficient $1 + \phi_1 - \lambda\phi_2$. Therefore, the series $Y_{1,t} - \lambda Y_{2,t}$ is $I(0)$, unit-root nonstationary, or an explosive series in the cases where $|1 + \phi_1 - \lambda\phi_2|$ is less than 1, equal to 1, and greater than 1, respectively.

If $\phi_1 - \lambda\phi_2 > 0$, then $1 + \phi_1 - \lambda\phi_2 > 1$ and $Y_{1,t} - \lambda Y_{2,t}$ is explosive. If $\phi_1 - \lambda\phi_2 = 0$, then $1 + \phi_1 - \lambda\phi_2 = 1$ and $Y_{1,t} - \lambda Y_{2,t}$ is a random walk. If $\phi_1 - \lambda\phi_2 < 0$, then $1 + \phi_1 - \lambda\phi_2 < 1$ and $Y_{1,t} - \lambda Y_{2,t}$ is stationary, unless $\phi_1 - \lambda\phi_2 < -2$ so that $1 + \phi_1 - \lambda\phi_2 \leq -1$.

The case $\phi_1 - \lambda\phi_2 \leq -2$ is "over-correction." The change in $Y_{1,t} - \lambda Y_{2,t}$ is in the correct direction but too large, so the series oscillates in sign but diverges to ∞ in magnitude.

Example 15.2. Simulation of error correction model

Model (15.2)–(15.3) was simulated with $\phi_1 = 0.5$, $\phi_2 = 0.55$, and $\lambda = 1$. A total of 5000 observations was simulated, but, for visual clarity, only every 10th observation is plotted in Figure 15.2. Neither $Y_{1,t}$ nor $Y_{2,t}$ is stationary, but $Y_{1,t} - \lambda Y_{2,t}$ is stationary. Notice how closely $Y_{1,t}$ and $Y_{2,t}$ track each other.

□

To see how to generalize error correction to more than two series, it is useful to rewrite equations (15.2) and 15.3) in vector form. Let $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t})^{\mathsf{T}}$ and $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{2,t})^{\mathsf{T}}$. Then

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{Y}_{t-1} + \boldsymbol{\epsilon}_t, \tag{15.5}$$

where

$$\boldsymbol{\alpha} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ -\lambda \end{pmatrix}, \tag{15.6}$$

so that $\boldsymbol{\beta}$ is the cointegration vector and $\boldsymbol{\alpha}$ specifies the speed of mean-reversion and is called the *loading matrix* or *adjustment matrix*.

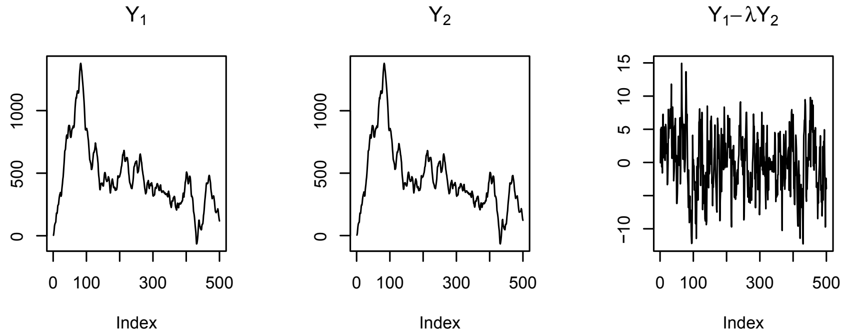

Fig. 15.2. *Simulation of an error correction model. 5000 observations were simulated but only every 10th is plotted.*

Model (15.5) also applies when there are d series so that $\boldsymbol{Y}_t$ and $\boldsymbol{\epsilon}_t$ d-dimensional. In this case $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are each full-rank $d \times r$ matrices for some $r \leq d$ which is the number of linearly independent cointegration vectors. The columns of $\boldsymbol{\beta}$ are the cointegration vectors.

Model (15.5) is a vector AR(1) [that is, VAR(1)] model but, for added flexibility, can be extended to a VAR(p) model, and there are several ways to do this. We will use the notation and the second of two forms of the VECM from the function `ca.jo` in `R`'s `urca` package. This VECM is

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta \boldsymbol{Y}_{t-p+1} + \boldsymbol{\Pi} \boldsymbol{Y}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\epsilon}_t, \quad (15.7)$$

where $\boldsymbol{\mu}$ is a mean vector, $\boldsymbol{D}_t$ is a vector of nonstochastic regressors, and

$$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}}. \quad (15.8)$$

As before, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are each full-rank $d \times r$ matrices and $\boldsymbol{\alpha}$ is called the loading matrix.

It is easy to show that the columns of $\boldsymbol{\beta}$ are the cointegration vectors. Since $\boldsymbol{Y}_t$ is $I(1)$, $\Delta \boldsymbol{Y}_t$ on the left-hand side of (15.7) is $I(0)$ and therefore $\boldsymbol{\Pi} \boldsymbol{Y}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{Y}_{t-1}$ on the right-hand side of (15.7) is also $I(0)$. It follows that each of the r components of $\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{Y}_{t-1}$ is $I(0)$.

Example 15.3. VECM test on bond yields

A VECM was fit to the bond yields using `R`'s `ca.jo` function. The output is below. The eigenvalues are used to test null hypotheses of the form H_0: $r \leq r_0$. The values of the test statistics and critical values (for 1%, 5%, and 10% level tests) are listed below the eigenvalues. The null hypothesis is rejected when the test statistic exceeds the critical level. In this case, regardless of whether one uses a 1%, 5%, or 10% level test, one accepts that r is less than or equal to 3 but rejects that r is less than or equal to 2, so one concludes that $r = 3$. Although five cointegration vectors are printed, only the first three would be meaningful. The cointegration vectors are the columns of the matrix labeled "Eigenvectors, normalised to first column." The cointegration vectors are determined only up to multiplication by a nonzero scalar and so can be normalized so that their first element is 1.

```
######################
# Johansen-Procedure #
######################

Test type: maximal eigenvalue statistic (lambda max),
with linear trend

Eigenvalues (lambda):
[1] 0.03436 0.02377 0.01470 0.00140 0.00055

Values of test statistic and critical values of test:

          test 10pct  5pct 1pct
r <= 4 |   2.59   6.5  8.18 11.6
r <= 3 |   6.62  12.9 14.90 19.2
r <= 2 |  69.77  18.9 21.07 25.8
r <= 1 | 113.36  24.8 27.14 32.1
r = 0  | 164.75  30.8 33.32 38.8

Eigenvectors, normalised to first column:
(These are the cointegration relations)

        X3mo.l2 X6mo.l2 X1yr.l2 X2yr.l2 X3yr.l2
X3mo.l2   1.000    1.00    1.00  1.0000   1.000
X6mo.l2  -1.951    2.46    1.07  0.0592   0.897
X1yr.l2   1.056   14.25   -3.95 -2.5433  -1.585
X2yr.l2   0.304  -46.53    3.51 -3.4774  -0.118
X3yr.l2  -0.412   30.12   -1.71  5.2322   1.938

Weights W:
(This is the loading matrix)

        X3mo.l2   X6mo.l2   X1yr.l2   X2yr.l2   X3yr.l2
X3mo.d -0.03441 -0.002440 -0.011528 -0.000178 -0.000104
X6mo.d  0.01596 -0.002090 -0.007066  0.000267 -0.000170
X1yr.d -0.00585 -0.001661 -0.001255  0.000358 -0.000289
X2yr.d  0.00585 -0.000579 -0.003673 -0.000072 -0.000412
X3yr.d  0.01208 -0.000985 -0.000217 -0.000431 -0.000407
```

□

15.3 Trading Strategies

As discussed previously, price series that are cointegrated can be used in *statistical arbitrage*. Unlike pure arbitrage, statistical arbitrage means an opportunity where a profit is only likely, not guaranteed. Pairs trading uses pairs of cointegrated asset prices and has been a popular statistical arbitrage technique. Pairs trading requires the trader to find cointegrated pairs of assets, to select from these the pairs that can be traded profitably after accounting for transaction costs, and finally to design the trading strategy which includes the buy and sell signals. A full discussion of statistical arbitrage is outside the scope of this book, but see Section 15.4 for further reading.

Although many firms have been very successful using statistical arbitrage, one should be mindful of the risks. One is model risk; the error-correction model may be incorrect. Even if the model is correct, one must use estimates based on past data and the parameters might change, perhaps rapidly. If statistical arbitrage opportunities exist, then it is possible that other traders have discovered them and their trading activity is one reason to expect parameters to change. Another risk is that one can go bankrupt before a stationary process reverts to its mean. This risk is especially large because firms engaging in statistical arbitrage are likely to be heavily leveraged. High leverage will magnify a small loss caused when a process diverges even farther from its mean before reverting. See Sections 2.5.2 and 15.6.3.

15.4 Bibliographic Notes

Alexander (2001), Enders (2004), and Hamilton (1994) contain useful discussions of cointegration. Pfaff (2006) is a good introduction to the analysis of cointegrated time series using `R`.

The MLEs and likelihood ratio tests of the parameters in (15.7) were developed by Johansen (1991, 1995) and Johansen and Juselius (1990).

The applications of cointegration theory in statistical arbitrage are discussed by Vidyamurthy (2004) and Alexander, Giblin, and Weddington (2001). Pole (2007) is a less technical introduction to statistical arbitrage.

15.5 References

Alexander, C. (2001) *Market Models: A Guide to Financial Data Analysis*, Wiley, Chichester.

Alexander, C., Giblin, I., and Weddington, W. III (2001) *Cointegration and Asset Allocation: A New Hedge Fund,* ISMA Discussion Centre Discussion Papers in Finance 2001–2003.

Enders, W. (2004) *Applied Econometric Time Series*, 2nd ed., Wiley, New York.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton University Press, Princeton, NJ.
Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, **59**, 1551-1580.
Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, New York.
Johansen, S., and Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration – With applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, **52**, 2, 169-210.
Pfaff, B. (2006) *Analysis of Integrated and Cointegrated Time Series with R*, Springer, New York.
Phillips, P. C. B., and Ouliaris, S. (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica*, **58**, 165–193.
Pole, A. (2007) *Statistical Arbitrage*, Wiley, Hoboken, NJ.
Vidyamurthy, G. (2004) *Pairs Trading*, Wiley, Hoboken, NJ.

15.6 R Lab

15.6.1 Cointegration Analysis of Midcap Prices

The data set `midcapD.ts` in the `fEcofin` package has daily returns on 20 midcap stocks in columns 2–21. Columns 1 and 22 contain the date and market returns, respectively. In this section, we will use returns on the first 10 stocks. To find the stock prices from the returns, we use the relationship

$$P_t = P_0 \exp(r_1 + \cdots + r_t),$$

where P_t and r_t are the price and log return at time t. The returns will be used as approximations to the log returns. The prices at time 0 are unknown, so we will use $P_0 = 1$ for each stock. This means that the price series we use will be off by multiplicative factors. This does not affect the number of cointegration vectors. If we find that there are cointegration relationships, then it would be necessary to get the price data to investigate trading strategies.

Johansen's cointegration analysis will be applied to the prices with the `ca.jo` function in the `urca` package. Run

```
library(fEcofin)
library(urca)
x = midcapD.ts[,2:11]
prices= exp(apply(x,2,cumsum))
options(digits=3)
summary(ca.jo(prices))
```

Problem 1 *How many cointegration vectors were found?*

15.6.2 Cointegration Analysis of Yields

This example is similar to Example 15.3 but uses different yield data. The data are in the `mk.zero2` data set in the `fEcofin` package. There are 55 maturities and they are in the vector `mk.maturity`. We will use only the first 10 yields. Run

```
library("fEcofin")
library(urca)
mk.maturity[2:11,]
summary(ca.jo(mk.zero2[,2:11]))
```

Problem 2 *What maturities are being used? Are they short-, medium-, or long-term, or a mixture of short- and long-term maturities?*

Problem 3 *How many cointegration vectors were found? Use 1% level tests.*

15.6.3 Simulation

In this section, you will run simulations similar to those in Section 2.5.2. The difference is that now the price process is mean-reverting.

Suppose a hedge fund owns a \$1,000,000 position in a portfolio and used \$50,000 of its own capital and \$950,000 in borrowed money for the purchase. If the value of the portfolio falls below \$950,000 at the end of any trading day, then the hedge fund must liquidate and repay the loan.

The portfolio was selected by cointegration analysis and its price is an AR(1) process,

$$(P_t - \mu) = \phi(P_{t-1} - \mu) + \epsilon_t,$$

where P_t is the price of the portfolio at the end of trading day t, $\mu =$ \$1,030,000, $\phi = 0.99$, and the standard deviation of ϵ_t is \$5000. The hedge fund knows that the price will eventually revert to \$1,030,000 (assuming that the model is correct and, of course, this is a big assumption). It has decided to liquidate its position on day t if $P_t \geq$ \$1,020,000. This will yield a profit of at least \$20,000. However, if the price falls below \$950,000, then it must liquidate and lose its entire \$50,000 investment plus the difference between \$950,000 and the price at liquidation.

In summary, the hedge fund will liquidate at the end of the first day such that the price is either above \$1,020,000 or below \$950,000. In the first case, it will achieve a profit of at least \$20,000 and in the second case it will suffer a loss of at least \$50,000. Presumably, the probability of a loss is small, and we will see how small by simulation.

Run a simulation experiment similar to the one in Section 2.5.2 to answer the following questions. Use 10,000 simulations.

Problem 4 *What is the expected profit?*

Problem 5 *What is the probability that the hedge fund will need to liquidate for a loss?*

Problem 6 *What is the expected waiting time until the portfolio is liquidated?*

Problem 7 *What is the expected yearly return on the \$50,000 investment?*

15.7 Exercises

1. Show that (15.4) implies that $Y_{1,t-1} - \lambda Y_{2,t-1}$ is an AR(1) process with coefficient $1 + \phi_1 - \lambda\phi_2$.
2. In (15.2) and (15.3) there are no constants, so that $Y_{1,t} - \lambda Y_{2,t}$ is a stationary process with mean zero. Introduce constants into (15.2) and (15.3) and show how they determine the mean of $Y_{1,t} - \lambda Y_{2,t}$.
3. Verify that in Example 15.2 $Y_{1,t} - \lambda Y_{2,t}$ is stationary.
4. Suppose that $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t})^{\mathsf{T}}$ is the bivariate AR(1) process in Example 15.2. Is $\boldsymbol{Y}_t$ stationary? (Hint: See Section 10.3.3.)