COINTEGRATION: A REVIEW

by

JIE ZHANG

B.A., Peking University, 2006

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts And Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2008

Approved by:

Major Professor
Shien-Shien Yang

Abstract

Many nonstationary univariate time series can be made stationary by appropriate differencing before ARMA models are fitted to the differenced series. However, when it comes to nonstationary vector time series, the situation is more complex. Since the dynamic of a multivariate time series is multidimensional, even if we can make each component stationary by appropriate differencing, the vector process of the differenced components may be still nonstationary. However, it is possible that the projections of a nonstationary vector time series in some directions may result in a stationary process. Engle and Granger(1987) formally demonstrated that it is possible for some linear combinations of the components of nonstationary vector time series to be stationary. They called this phenomenon Co-Integration.

This concept of cointegration turned out to be extremely important in the modeling and analysis of non-stationary time series in economics. Although economic variables individually may exhibit disequilibrium behaviors, often time, due to economic forces, these disequilibrium economic variables corporately form a dynamic equilibrium relationship. Specifically, certain linear combinations of nonstationary time series may appear to be stationary. Engle and Granger developed statistical method for detecting and estimating this equilibrium relationship. They also proposed the so called error correction model to model Co-Integrated vector time series.

In this report, I give a detail review on the concept of cointegration, the 2-step estimation procedure for the error correction models, and the 7 types of tests for testing cointegration. Since the test statistics for testing cointegration do not follow any known distribution, critical values were obtained based on two models by Engle and Granger. Augmented Dickey-Fuller and Dickey-Fuller tests were recommended as it is believed that their distributions are independent of the under lying process model. The critical values table presented in their paper is widely used in testing cointegration. In this report, we’ll construct tables of critical values based on different models and compare them with those obtained by Engle and Granger. Also, to demonstrate the practical usage of cointegration, applications to currency exchange rates and US stock and Asian stock indexes are presented as illustrative examples.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The accomplishment of this paper is due to the effort of my major advisor: Dr. Yang, together with my committee members: Dr. Song and Dr. Nelson. I would like to thank them for their advices and supports.

Special thanks to Mom and Dad. I won’t be here if it were not for them.

CHAPTER 1 - Introduction

Many nonstationary univariate time series can be made stationary by appropriate differencing before ARMA models are fitted to the differenced series. However, when it comes to nonstationary vector time series, the situation is more complex. Since the dynamic of a multivariate time series is multidimensional, even if we can make each component stationary by appropriate differencing, the vector process of the differenced components may be still nonstationary. However, it is possible that the projections of a nonstationary vector time series in some directions may result in a stationary process. Engle and Granger(1987) formally demonstrated that it is possible for some linear combinations of the components of nonstationary vector time series to be stationary. They called this phenomenon Co-Integration.

This concept of cointegration turned out to be extremely important in the modeling and analysis of non-stationary time series in economics. Although economic variables individually may exhibit disequilibrium behaviors, often time, due to economic forces, these disequilibrium economic variables corporately form a dynamic equilibrium relationship. Specifically, certain linear combinations of nonstationary time series may appear to be stationary. Engle and Granger developed statistical method for detecting and estimating this equilibrium relationship. They also proposed the so called error correction model to model Co-Integrated vector time series.

This paper gives a detail review of the concept of cointegration.

The second chapter briefly introduces basic notation, representations of vector time series and definition of stationarity. Chapter 3 states the definition of cointegration, several representations of cointegrated vector processes and the two-step method for estimating the cointegrating system proposed by Engle and Granger (1987). Chapter 4 discusses 7 types of cointegration tests for bivariate $CI(1,1)$ case and provides critical values based on several null hypothesis generating models. Two applications of cointegration on finance data are presented in Chapter 5 as examples. Codes for simulations and examples are provided in the appendix.

CHAPTER 2 - Vector Time Series

Time series data in many empirical studies, especially those involved in economics and finances, consist of observations from several variables. The interrelation and dynamic among all the variables are of great interest. For example, as economic globalization and internet communication accelerating the integration of world financial markets in recent years, financial markets are more and more dependent on each other than ever before. Hence, to understand the dynamic structure of the global finance, one need to consider several financial and economical variables simultaneously representing the behaviors of different markets. In such cases, multivariate time series models are used to describe interrelationships among several time series variables.

This chapter is organized as following. Section1 defines weak stationarity and correlation structures of a vector time series. Section 2 discusses two widely used models of multivariate time series: Moving average and Autoregressive representations. Section 3 introduces the definitions of nonstationarity and some of the studies on multivariate time series.

Section 2.1 Weak Stationarity

Consider a m-dimensional time series $\mathbf{Z_t} = [Z_{1,t}, Z_{2,t}, ..., Z_{m,t}]'$, $t = 0, \pm 1, \pm 2, \ldots$. A vector time series $\mathbf{Z_t}$ is **weakly stationary** if its first and second moments are time-invariant. In particular, the mean vector of a weakly stationary series is constant over time and the cross-covariance between $Z_{i,t}$ and $Z_{j,s}$, for all i=1,2,…m and j=1,2,…m, are functions only of the absolute value of the time difference |s-t|

For a weakly stationary time series $\mathbf{Z_t}$, we define its mean vector to be $E(\mathbf{Z_t}) = \boldsymbol{\mu}$ and the lag-k covariance matrix as:

$$\boldsymbol{\Gamma}(k) = \operatorname{cov}\{\mathbf{Z}_t, \mathbf{Z}_{t+k}\} = E[(\mathbf{Z}_t - \boldsymbol{\mu})(\mathbf{Z}_{t+k} - \boldsymbol{\mu})']$$

$$= \begin{pmatrix} \gamma_{11}(k) & \ldots & \gamma_{1m}(k) \\ \vdots & \ddots & \vdots \\ \gamma_{m1}(k) & \cdots & \gamma_{mm}(k) \end{pmatrix},$$

where

$$\gamma_{ij}(k) = E(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j),$$

for $k = 0, \pm 1, \pm 2..., i = 1, 2, ..., m,$ and $j = 1, 2, ..., m$. As a function of k, $\mathbf{\Gamma}(k)$ is also called the covariance matrix function for the vector process $\mathbf{Z_t}$. $\mathbf{\Gamma}(0)$ is called the contemporaneous variance-covariance matrix of the vector process $\mathbf{Z_t}$. And $\gamma_{ii}(k)$ is the auto-covariance function for the ith component process $Z_{i,t}$; $\gamma_{ij}(k)$ is the cross-covariance function between $Z_{i,t}$ and $Z_{j,t}$ for $i \neq j$.

The correlation matrix function is defined by

$$\boldsymbol{\rho}(k) = \mathbf{D}^{-1/2}\mathbf{\Gamma}(k)\mathbf{D}^{-1/2} = [\rho_{ij}(k)]$$

where $\mathbf{D}$ is the diagonal matrix with $\gamma_{ii}(0)$, (i=1,2,…,m) as its diagonal elements. The ith diagonal element of $\boldsymbol{\rho}(k)$, $\rho_{ii}(k)$ is the autocorrelation function for the ith component series $Z_{i,t}$ whereas the off-diagonal element $\rho_{ij}(k)$, $i \neq j$, represents the cross-correlation function between $Z_{i,t}$ and $Z_{j,t}$.

Like the univariate autocovariance and autocorrelation functions, the covariance and correlation matrix functions are also semi-positive definite in the following sense

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i'\mathbf{\Gamma}(t_i - t_j)\boldsymbol{\alpha}_j \geq 0$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i'\boldsymbol{\rho}(t_i - t_j)\boldsymbol{\alpha}_j \geq 0$$

for any set of time points $t_1, t_2, \ldots, t_n$ and any set of real vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_n$.

Since $\gamma_{ij}(k) = E(Z_{i,t} - \mu_i)(Z_{j,t+k} - \mu_j) = E(Z_{j,t+k} - \mu_j)(Z_{i,t} - \mu_i) = \gamma_{ji}(-k)$, $\Gamma(k)$ and $\rho(k)$ have the following symmetry property:

$$\begin{cases} \mathbf{\Gamma}(k) = \mathbf{\Gamma}'(-k) \\ \boldsymbol{\rho}(k) = \boldsymbol{\rho}'(-k) \end{cases}$$

Sometimes, the covariance and correlation matrix functions are also called autocovariance and autocorrelation functions.

Section 2.2 Some Vector Time Series Models

An m-dimensional vector process $\mathbf{Z}_t$ is said to be a purely nondeterministic vector process if it can be written as a weighted sum of sequence of m-dimensional white noise random process. Namely,

$$\begin{aligned}\mathbf{Z}_t &= \boldsymbol{\mu} + \mathbf{a}_t + \boldsymbol{\Psi}_1\mathbf{a}_{t-1} + \boldsymbol{\Psi}_2\mathbf{a}_{t-2} + \cdots \\ &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Psi}_s \mathbf{a}_{t-s}\end{aligned}$$

where $\boldsymbol{\Psi}_0 = \mathbf{I}$ is the mxm identity matrix, the $\boldsymbol{\Psi}_j{}'s$ are mxm coefficient matrices, and the $\mathbf{a}_t$ ’s are m-dimensional white noise random vectors with zero mean and covariance matrix

$$E[\mathbf{a}_t\mathbf{a}'_{t+k}] = \begin{cases}\boldsymbol{\Sigma}, & \text{if k=0} \\ 0, & \text{if } k \neq 0\end{cases}$$

with $\boldsymbol{\Sigma}$ being a mxm symmetric positive definite matrix. Hence, even though the components of $\mathbf{a}_t$ at different times are uncorrelated, they might be contemporaneously correlated. Using the backshift operator B, and with $\dot{\mathbf{Z}}_t = \mathbf{Z}_t - \boldsymbol{\mu}$, the equivalent representation of the above model can be written as

$$\dot{\mathbf{Z}}_\mathbf{t} = \boldsymbol{\Psi}(B)\mathbf{a}_t = \sum_{\mathbf{s=0}}^{\infty} \boldsymbol{\Psi}_s B^s \mathbf{a}_\mathbf{t}\,.$$

This presentation is known as the ***vector moving average*** or ***Wold representation***.

Let $\boldsymbol{\Psi}_s = [\psi_{ij,s}]$ and $\boldsymbol{\Psi}(B) = [\psi_{ij}(B)]$ where $\psi_{ij}(B) = \sum_{s=0}^{\infty} \psi_{ij,s} B^s$. If the coefficient matrices $\boldsymbol{\Psi}_s$ is square summable, in the sense that each of the mxm sequences $\psi_{ij,s}$ is square summable, i.e., $\sum_{s=0}^{\infty} \psi^2{}_{ij,s} < \infty$ for i=1,2,…,m, j=1,2,…,m, then we say the vector process is ***stationary***.

Another useful representation of a multivariate time series is that apart from a white noise process $\mathbf{a}_t$, $\mathbf{Z}_t$ is a linear function of its past:

$$\begin{aligned}\dot{\mathbf{Z}}_\mathbf{t} &= \boldsymbol{\Pi}_1\dot{\mathbf{Z}}_\mathbf{t-1} + \boldsymbol{\Pi}_2\dot{\mathbf{Z}}_\mathbf{t-2} + \cdots + \mathbf{a}_\mathbf{t} \\ &= \sum_{\mathbf{s=1}}^{\infty} \boldsymbol{\Pi}_s \dot{\mathbf{Z}}_\mathbf{t-s} + \mathbf{a}_\mathbf{t}\end{aligned}$$

In terms of backshift operator,

$$\boldsymbol{\Pi}(B)\dot{\mathbf{Z}}_t = \mathbf{a}_t$$

where

$$\mathbf{\Pi}(B) = \mathbf{I} - \sum_{\mathbf{s=1}}^{\infty} \mathbf{\Pi_s} B^s$$

and $\mathbf{\Pi_s}$ are $m \times m$ autoregressive coefficient matrices. The above representation is called ***vector autoregressive (VAR) representation***.

Combined the two representation, the widely used ***vector autoregressive moving average*** VARMA(p,q) process expressed in backshift operator is of the form

$$\mathbf{\Phi}_p(B)\dot{\mathbf{Z}}_t = \mathbf{\Theta}_q(B)\mathbf{a}_t$$

where

$$\mathbf{\Phi}_p(B) = \mathbf{\Phi}_0 - \mathbf{\Phi}_1 B - \mathbf{\Phi}_2 B^2 - \cdots - \mathbf{\Phi}_p B^p$$

and

$$\mathbf{\Theta}_q(B) = \mathbf{\Theta}_0 - \mathbf{\Theta}_1 B - \mathbf{\Theta}_2 B^2 - \cdots - \mathbf{\Theta}_q B^q$$

are the autoregressive and moving average matrix polynomials of orders p and q respectively. We assume that the two matrix polynomials have no left common factors; otherwise, we can simplify the model. When $\mathbf{\Sigma}$ (the covariance matrix of $\mathbf{a}_t$) is positive definite, without loss of generality we can also assume that $\mathbf{\Phi}_0 = \mathbf{\Theta}_0 = \mathbf{I}$, an mxm identity matrix. By taking p=0 or q=0, it is easily seen that moving average and autoregressive processes are just special cases of ARMA representation.

The process is stationary if the zeros of the determinantal polynomial $\left|\mathbf{\Phi}_p(B)\right|$ are outside the unit circle. In this case, writing

$$\mathbf{\Psi}(B) = [\mathbf{\Phi}_p(B)]^{-1}\mathbf{\Theta}_q(B)$$

then the equivalent moving average representation is

$$\dot{\mathbf{Z}}_t = \mathbf{\Psi}(B)\mathbf{a}_t = \sum_{s=0}^{\infty} \mathbf{\Psi}_s B^s$$

and the sequence $\mathbf{\Psi}_s$ is square summable.

When a vector time series is stationary, and a model is identified, the fitting of the model can be obtained by maximizing the likelihood function if we assume the vector time series is a Gaussian process.

However, when the time series is not stationary, maximum likelihood procedure is not directly applicable.

Section 2.3 Nonstationarity

In the analysis of time series, it is not unusual to observe series that exhibit nonstationary behavior. One useful and most frequently used way to reduce nonstationary univariate time series to stationary series is by appropriate differencing. For example, in univariate time series, a nonstationary series Z_t can be reduced to a stationary series $(1-B^s)^d Z_t$ for an appropriate choice of d>0 and $s > 0$, so that we can write

$$\phi_p(B)(1-B^s)^d Z_t = \theta_q(B)a_t$$

with $\phi_p(B)$ a stationary AR operator. A natural, an extension to the vector process is

$$\mathbf{\Phi}_p(B)(\mathbf{I}-\mathbf{I}B^s)^d \mathbf{Z}_t = \mathbf{\Theta}_q(B)\mathbf{a}_t$$

i.e.,

$$\mathbf{\Phi}_p(B)(1-B^s)^d \mathbf{Z}_t = \mathbf{\Theta}_q(B)\mathbf{a}_t$$

This extension implies that all component series are differenced the same number of times, which is unnecessary and undesirable in most cases. To be more flexible, we assume that $\mathbf{Z}_t$ can be reduced to stationary vector series by applying a differencing operator $\mathbf{D}(B)$, where

$$\mathbf{D}(B) = \begin{bmatrix} (1-B^{s_1})^{d_1} & 0 & \cdots 0 & 0 \\ 0 & (1-B^{s_2})^{d_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (1-B^{s_m})^{d_m} \end{bmatrix}$$

and $(d_1, d_2, \ldots, d_m)$&$(s_1, s_2, \ldots, s_m)$ are two sets of nonnegative integers such that we have a nonstationary vector ARMA model for $\mathbf{Z}_t$

$$\mathbf{\Phi}_p(B)\mathbf{D}(B)\mathbf{Z}_t = \mathbf{\Theta}_q(B)\mathbf{a}_t$$

for which the zeros of $|\mathbf{\Phi}_p(B)|$ are outside the unit circle.

However, compared with univariate case, differencing on vector time series is much more complicated. Over differencing may lead to complications in model fitting. And Box and Tiao (1977) shows that when the orders of differencing for each component series are the same, it may lead to a noninvertible representation. Hence, one should be particularly careful when handle the nonstationary vector processes by differencing. Box and Tiao (1977) also points out that when zeros of $|\mathbf{\Phi}_p(B)|$ approach values on the unit circle, a canonical transformation can decompose $\mathbf{Z}_t$ into two parts, one of which follows a stationary autoregressive process, while the other part approaches nonstationarity. They suggested that, in analyzing multiple time series, "it is useful to entertain the possibility that the dynamic pattern in the data may be due to a small subset of nearly nonstationary components and that there may exist stable contemporaneous linear relationships among the variables." Hence, differencing of the original series could lead to complications in the analysis. Especially, when a linear combination of the component series is stationary, a model purely based on differences may not even exist. For example, suppose we have a bivariate model

$$x_{1t} = x_{1(t-1)} + a_{1t},\ x_{2t} = \beta x_{1t} + a_{2t}$$

Each series individually is nonstationary, but the linear combination of the two components $x_{2t} - \beta x_{1t}$ is stationary. By differencing the two series, we get:

$$\begin{aligned} w_{1t} &= (1-B)x_{1t} = a_{1t}, \\ w_{2t} &= (1-B)x_{2t} = \beta a_{1t} + a_{2t} - a_{2(t-1)} \end{aligned}$$

i.e.

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

The differenced series can not be expressed in the form of bivariate stationary autoregressive process any more, making the analysis more complicated. Hence, to identify and estimate the possible stationary linear combination of the components and build an estimable model for this type of time series are of great importance in multivariate time series analysis.
(Find an example, each component is stationary but jointly nonstationary)

CHAPTER 3 - Cointegration

As stated in the last chapter, most statistical theory applied in building, estimating and testing time series models are based on the assumption that the time series in the models are stationary. Statistical inference associated with a time series process is not valid if the assumption of stationarity is violated. However, nonstationarity is a common property to many time series. Especially in macroeconomic and financial processes, often time a process has no clear tendency to return to a constant value or a linear trend. By appropriate differencing, one can achieve stationary components but it might complicate the structure of the time series. Fortunately, although individual time series can wander extensively, some subset of these series may move in a pattern so that they do not drift too far apart from each other. Such phenomenon can be found in financial and economic time series data—for examples, indexes of different stock markets or exchange rates among different currencies. We consider such phenomenon as existence of an equilibrium relationship between the nonstationary time series. To describe this phenomenon, Clive Granger first introduced the concept of cointegration, which was thought of as a great breakthrough and has changed the way empirical models of macroeconomic relationships are formulated today.

As a review of Engle and Granger (1987)'s work, the first section of this chapter introduces the definition of cointegration. Section 2 presents the two equivalent models of cointegrated time series: Granger's representation and error correction model. Section 3 provides methods for estimating cointegrated systems.

Section 3.1 Definition of Cointegration

It is well known from Wold's theorem that a single stationary time series with no deterministic components has an infinite moving average representation. If in addition, it is invertible, then it can be approximated by a finite autoregressive moving average process. Many nonstationary time series can be made stationary by appropriate differencing. The following definition formally defines such a class of nonstationary time series.

Definition: A series with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted by $x_t \sim I(d)$.

Under this notion, $x_t \sim I(0)$ is stationary while $x_t \sim I(1)$ is nonstationary but has a stationary change. There are substantial differences in behavior between a series that is $I(0)$ and another which is $I(1)$. Suppose $x_t \sim I(0)$ with zero mean, then (i) the variance of x_t is a finite constant; (ii) an innovation has only a temporary effect on the value of x_t; (iii) the autocorrelations, ρ_k, decrease rapidly in magnitude as k increases, so that the infinite sum of them is finite. Whereas, if $x_t \sim I(1)$ with $x_0 = 0$, then (i) variance of x_t goes to infinity as t goes to infinity; (ii) its innovation has a permanent effect on the value of x_t, since x_t is the sum of all previous changes; (iii) the theoretical autocorrelations, $\rho_k \to 1$ for all k as $t \to \infty$. More discussion can be found in Feller (1968) or Granger and Newbold (1977).

Due to the relative sizes of the variances, it is always true that the sum of an $I(0)$ and $I(1)$ will be $I(1)$. Generally, if a and b are constants, $b \neq 0$, then $a + bx_t$ is $I(d)$ if x_t is $I(d)$; However, if x_t and y_t are both $I(d)$, then it is possible that the linear combination $z_t = x_t - ay_t$ will be $I(d-b)$, $b > 0$. Consider the case when $d = b = 1$, so that x_t and y_t are both $I(1)$ with dominant long run components, but their linear combination z_t is $I(0)$, a stationary series. This is a special constraint on the long-run components of the two series. However, it is worth noticing that it is not generally true that there exists such an a that makes $z_t \sim I(0)$. To formalize the ideas above, the following definition from Granger(1981) and Granger and Weiss(1983) is introduced.

Definition: The components of the vector $\mathbf{x_t}$ are said to be ***co-integrated of order*** d, b, denoted $\mathbf{x_t} \sim CI(d,b)$, if (i) all components of $\mathbf{x_t}$ are $I(d)$; (ii) there exists a vector $\boldsymbol{\alpha}(\neq 0)$ so that $z_t = \boldsymbol{\alpha}'\mathbf{x_t} \sim I(d-b),\ b > 0$. The vector $\boldsymbol{\alpha}$ is called a ***co-integrating vector.***

As an illustration, consider the vector time series in section 2.3,

$$x_{1t} = x_{1(t-1)} + a_{1t},\ x_{2t} = \beta x_{1t} + a_{2t}.$$

Clearly, each component is an $I(1)$ process, since they become stationary after first differencing:

$$\begin{aligned} w_{1t} &= (1-B)x_{1t} = a_{1t}, \\ w_{2t} &= (1-B)x_{2t} = \beta a_{1t} + a_{2t} - a_{2(t-1)}. \end{aligned}$$

However, with $\boldsymbol{\alpha}' = [\beta, -1]$, the linear combination

$$\alpha' x_t = [\beta, -1]\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \beta x_{1t} - x_{2t} = \beta x_{1t} - \beta x_{1t-1} - a_{2t} = \beta a_{1t} - a_{2t}$$

is $I(0)$. Hence, the vector time series $\mathbf{x}_t$ is cointegrated with a cointegration vector $\boldsymbol{\alpha}' = [\beta, -1]$.

Sometimes, co-integration vector is also called the long-run parameter. It is clearly not unique. Because if $\boldsymbol{\alpha}'\mathbf{x}_t$ is stationary, then so too is $c\boldsymbol{\alpha}'\mathbf{Z}_t$ for any nonzero constant c. Hence $c\boldsymbol{\alpha}$ is also a cointegrating vector. If $\mathbf{x_t}$ is a vector of economic variables, then they are said to be in equilibrium when the following linear constraint is satisfied.

$$\boldsymbol{\alpha}'\mathbf{x}_t = 0$$

Of course, in reality, the equilibrium holds only approximately in the sense that

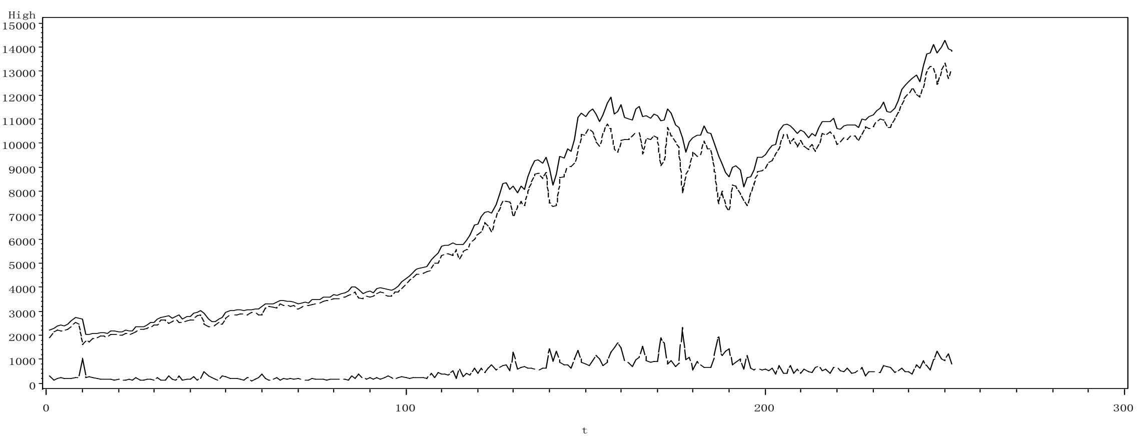
$$z_t = \boldsymbol{\alpha}'\mathbf{x}_t$$

is a $I(0)$ process, where z_t is called the equilibrium error.

Concentrating on the bivariate time series and d=1, b=1 case, cointegration would mean that if the components of vector time series $\mathbf{x_t}$ were all $I(1)$, then the equilibrium error would be $I(0)$, so that z_t will rarely drift far from zero if it has zero mean and will cross zero line often. It means that the equilibrium or at least a close approximation will occur often; whereas if $\mathbf{x_t}$ is not cointegrated, then for any vector $\boldsymbol{\alpha} \neq 0$, $z_t = \boldsymbol{\alpha}'\mathbf{x}_t$ will always wander widely and equilibrium would be rarely reached, which suggests that in this case equilibrium concept is not applicable.

The phenomenon of cointegration can be found in many economic studies. For example, as shown in Figure 3.1, the monthly highest quotation (the solid line) and lowest quotation (the dotted line) of Dow Jones industrial average are both individually nonstationary. However, the difference of the two series (the discontinuous line at the bottom) is $I(0)$, which indicates that although each series can wander wildly, they can not drift too far apart from each other.

Figure 3.1 Monthly high and low quotes of Dow Jones industrial average



More generally, if the vector series $\mathbf{x_t}$ contains p components, each being $I(1)$, then it is possible for several equilibrium relations to govern the joint behavior of the components of $\mathbf{x_t}$. So there may be k (<p) linearly independent cointegration vectors $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, ..., \boldsymbol{\alpha}_k$ such that $\boldsymbol{\alpha}'\mathbf{x}_t$ is a stationary (kx1) vector process, where

$$\boldsymbol{\alpha}' = \begin{bmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \\ \vdots \\ \boldsymbol{\alpha}_k' \end{bmatrix}$$

$\boldsymbol{\alpha}'$ is called the cointegrating matrix. If $\boldsymbol{\alpha}'$ is a cointegrating matrix, then for any qxk matrix $\mathbf{C}$, $\mathbf{C}\boldsymbol{\alpha}'$ is also a cointegrating matrix. Hence, $\boldsymbol{\alpha}$ is not unique. If for any other (1xp) vector $\mathbf{b}'$that is linearly independent of the rows of $\boldsymbol{\alpha}'$, we have that $\mathbf{b}'\mathbf{x}_t$ is nonstationary, then $\mathbf{x}_t$ is said to be ***cointegrated of rank k***. The vectors $\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', ..., \boldsymbol{\alpha}_k'$ form a basis for the space of the cointegrating vectors which is called ***cointegration space***.

Section 3.2 MA, AR Representations and Error Correction Model

Suppose that each component of $\mathbf{x}_t$ is $I(1)$, then without loss of generality we can assume that the change in each component is a zero mean purely nondeterministic stationary stochastic process, since any known deterministic components can be subtracted before the analysis is begun. It follows that there will always exist a multivariate Wold representation:

$$\Delta\mathbf{x_t} = (1-B)\mathbf{x}_t = \mathbf{\Psi}(B)\mathbf{a}_t$$

Where $\mathbf{\Psi}(B) = \sum_{j=0}^{\infty} \mathbf{\Psi}_j B^j$, $\mathbf{\Psi}(0) = \mathbf{I}$ and the coefficient matrices $\mathbf{\Psi}_j$ are absolutely summable since $(1-B)\mathbf{x}_t$ is stationary. $\mathbf{a}_t$ is the vector white noise process with mean 0 and covariance matrix

$$\begin{aligned} E(\mathbf{a}_s\mathbf{a}'_t) &= 0, \ t \neq s \\ &= \sigma^2, t = s \end{aligned}$$

so that only contemporaneous correlations can occur.

The moving average polynomial can be expressed as

$$
\begin{aligned}
\boldsymbol{\Psi}(B) &= \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j B^j = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j (B^j - 1) \\
&= \boldsymbol{\Psi}(1) + (1-B)(-\sum_{j=0}^{\infty} \boldsymbol{\Psi}_j (\sum_{i=0}^{j-1} B^i)) \\
&= \boldsymbol{\Psi}(1) + (1-B)\sum_{j=0}^{\infty} (-\sum_{i=1}^{\infty} \boldsymbol{\Psi}_{i+j}) B^j \\
&= \boldsymbol{\Psi}(1) + (1-B)\boldsymbol{\Psi}^*(B)
\end{aligned}
$$

Where $\boldsymbol{\Psi}^*(B) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}^*{}_j B^j$ and $\boldsymbol{\Psi}^*{}_j = -\sum_{i=1}^{\infty} \boldsymbol{\Psi}_{i+j}$. If $\boldsymbol{\Psi}(B)$ is of finite order, then $\boldsymbol{\Psi}^*(B)$ will be of finite order. If $\boldsymbol{\Psi}^*(1)$ is identically zero, then a similar expression involving $(1-B)^2$ can be defined. Based on the expression, $\mathbf{x_t}$ can be written in the form

$$
\begin{aligned}
\mathbf{x_t} - \mathbf{x}_0 &= \Delta\mathbf{x}_1 + \Delta\mathbf{x}_2 + \cdots + \Delta\mathbf{x}_t \\
&= \boldsymbol{\Psi}(B)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) \\
&= [\boldsymbol{\Psi}(1) + (1-B)\boldsymbol{\Psi}^*(B)](\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) \\
&= \boldsymbol{\Psi}(1)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) + (1-B)\boldsymbol{\Psi}^*(B)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t)
\end{aligned}
$$

Denote $\mathbf{y}_t = \boldsymbol{\Psi}^*(B)\mathbf{a}_t$, then

$$
\begin{aligned}
(1-B)\boldsymbol{\Psi}^*(B)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) &= (1-B)(\mathbf{y}_1 + \mathbf{y}_2 + \cdots \mathbf{y}_t) \\
&= \Delta\mathbf{y}_1 + \Delta\mathbf{y}_2 + \cdots + \Delta\mathbf{y}_t \\
&= \mathbf{y}_t - \mathbf{y}_0
\end{aligned}
$$

Hence

$$\mathbf{x_t} = \mathbf{x}_0 + \boldsymbol{\Psi}(1)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) + \mathbf{y}_t - \mathbf{y}_0$$

and

$$\boldsymbol{\alpha}'\mathbf{x_t} = \boldsymbol{\alpha}'(\mathbf{x}_0 - \mathbf{y}_0) + \boldsymbol{\alpha}'\boldsymbol{\Psi}(1)(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t) + \boldsymbol{\alpha}'\mathbf{y}_t$$

Obviously, $\mathbf{b}'(\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_t)$ is not stationary for any nonzero (1xp) vector $\mathbf{b}'$. Therefore, $\boldsymbol{\alpha}'\mathbf{x_t}$ will be stationary if and only if

$$\boldsymbol{\alpha}'\boldsymbol{\Psi}(1) = \mathbf{0}$$

This indicates that a cointegration matrix is perpendicular to $\boldsymbol{\Psi}(1)$. Thus the cointegration space spanned by the rows of $\boldsymbol{\alpha}'$ is a complement space of the column space of $\boldsymbol{\Psi}(1)$. The determinant $|\boldsymbol{\Psi}(B)| = 0$ at $B = 1$; hence the process is not invertible and we can never invert the MA representation $\Delta\mathbf{x_t} = \boldsymbol{\Psi}(B)\mathbf{a}_t$ to represent a cointegrated process with a vector AR form in terms of $\Delta\mathbf{x_t}$. The vector AR representation of a cointegrated process must be in terms of $\mathbf{x}_t$ directly.

The representation $\Delta\mathbf{x_t} = \mathbf{\Psi}(B)\mathbf{a}_t$ and the above restriction together is the ***MA representation*** of a cointegrated vector process.

Suppose that $\mathbf{x}_t$ is nonstationary and can be represented as $AR(p)$ model

$$\mathbf{\Phi_p}(B)\mathbf{x}_t = \mathbf{a}_t$$

such that $\left|\mathbf{\Phi_p}(B)\right| = 0$ contains some unit roots, where $\mathbf{\Phi_p}(B) = \mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p$. Multiply $(1-B)$ on both sides,

$$(1-B)\mathbf{\Phi_p}(B)\mathbf{x}_t = (1-B)\mathbf{a}_t \,.$$

If each component of $\mathbf{x_t}$ is $\boldsymbol{I}(1)$, then, the MA representation $(1-B)\mathbf{x}_t = \mathbf{\Psi}(B)\mathbf{a}_t$ can be transformed to

$$(1-B)\mathbf{\Phi_p}(B)\mathbf{x}_t = \mathbf{\Phi_p}(B)\mathbf{\Psi}(B)\mathbf{a}_t \,.$$

Comparing the above two, we have

$$(1-B)\mathbf{a}_t = \mathbf{\Phi_p}(B)\mathbf{\Psi}(B)\mathbf{a}_t$$

holds for any $\mathbf{a}_t$. It implies that

$$(1-B)\mathbf{I} = \mathbf{\Phi_p}(B)\mathbf{\Psi}(B)$$

for any B. Hence, if we take $B = 1$,

$$\mathbf{\Phi_p}(1)\mathbf{\Psi}(1) = \mathbf{0} \,.$$

$\mathbf{\Phi_p}(1)$ is perpendicular of $\mathbf{\Psi}(1)$, so it must belong to the cointegration space spanned by the rows of $\boldsymbol{\alpha}'$. That indicates

$$\mathbf{\Phi_p}(1) = \boldsymbol{\gamma}\boldsymbol{\alpha}'$$

for some (pxk) matrix $\boldsymbol{\gamma}$. The model $\mathbf{\Phi_p}(B)\mathbf{x}_t = \mathbf{a}_t$ and the above restriction together is the ***AR presentation*** of a cointegrated vector process.

Notice in the AR representation,

$$\mathbf{\Phi_p}(B) = \mathbf{I} - \mathbf{\Phi}_1 B - \cdots - \mathbf{\Phi}_p B^p = (\mathbf{I} - \boldsymbol{\lambda} B) - (\mathbf{\Phi}^*{}_1 B + \cdots + \mathbf{\Phi}^*{}_{p-1} B^{p-1})(1-B),$$

where $\boldsymbol{\lambda} = \mathbf{\Phi}_1 + \cdots + \mathbf{\Phi}_p$ and $\mathbf{\Phi}^*{}_j = -(\mathbf{\Phi}_{j+1} + \cdots + \mathbf{\Phi}_p)$ for $j = 1, 2, \ldots, p-1$. Hence the AR representation $\mathbf{\Phi_p}(B)\mathbf{x}_t = \mathbf{a}_t$ can be written as

$$(\mathbf{I}-\boldsymbol{\lambda}B)\mathbf{x}_t-(\boldsymbol{\Phi}^*{}_1B+\cdots+\boldsymbol{\Phi}^*{}_{p-1}B^{p-1})\Delta\mathbf{x}_t=\mathbf{a}_t$$

or

$$\mathbf{x}_t=\boldsymbol{\lambda}\mathbf{x}_{t-1}+\boldsymbol{\Phi}^*{}_1\Delta\mathbf{x}_{t-1}+\cdots+\boldsymbol{\Phi}^*{}_{p-1}\Delta\mathbf{x}_{t-p+1}+\mathbf{a}_t\,.$$

Subtract $\mathbf{x}_{t-1}$ on both sides, then

$$\Delta\mathbf{x}_t=\boldsymbol{\delta}\mathbf{x}_{t-1}+\boldsymbol{\Phi}^*{}_1\Delta\mathbf{x}_{t-1}+\cdots+\boldsymbol{\Phi}^*{}_{p-1}\Delta\mathbf{x}_{t-p+1}+\mathbf{a}_t$$

where $\boldsymbol{\delta}=\boldsymbol{\lambda}-\mathbf{I}=-\boldsymbol{\Phi}_p(1)=-\boldsymbol{\gamma}\boldsymbol{\alpha}'$ based on the restriction in AR representation. Therefore,

$$\Delta\mathbf{x}_t=-\boldsymbol{\gamma}\mathbf{z}_{t-1}+\boldsymbol{\Phi}^*{}_1\Delta\mathbf{x}_{t-1}+\cdots+\boldsymbol{\Phi}^*{}_{p-1}\Delta\mathbf{x}_{t-p+1}+\mathbf{a}_t$$

for some (pxk) matrix $\boldsymbol{\gamma}$, where $\mathbf{z}_{t-1}=\boldsymbol{\alpha}'\mathbf{x}_{t-1}$ is a (kx1) stationary process. This representation implies that the differenced series $\Delta\mathbf{x}_t$ of a cointegrated process $\mathbf{x}_t$ can not be described using only the values of its own lagged differences. The model must include an "error correction" term, $\boldsymbol{\gamma}\mathbf{z}_{t-1}=\boldsymbol{\gamma}\boldsymbol{\alpha}'\mathbf{x}_{t-1}$. Consider the relation $\Delta\mathbf{x}_t$ in terms of its own past lagged values as a long-run equilibrium, then the term $\mathbf{z}_{t-1}$ can be taken as an error from the equilibrium and the coefficient matrix $\boldsymbol{\gamma}$ is an adjustment for this error. Writing the above representation in an AR(p) form, we have the definition of error correction representation as below.

Definition: A vector time series $\mathbf{x}_t$ has an error correction representation if it can be expressed as:

$$\boldsymbol{\Phi}(B)(1-B)\mathbf{x}_t=-\gamma z_{t-1}+\mathbf{a}_t$$

where $\mathbf{a}_t$ is a stationary multivariate disturbance, with $\boldsymbol{\Phi}(0)=\mathbf{I}$, $\boldsymbol{\Phi}(1)$ has all elements finite, $z_t=\boldsymbol{\alpha}'\mathbf{x}_t$ and $\boldsymbol{\gamma}\neq 0$.

The error correction representation was first proposed by Davidson et al.(1978) and has been used widely in economic studies. For a two variable vector process, a typical error correction model would relate the change in one variable to past equilibrium errors, as well as to past changes in both variables.

The relationship between error correction models and co-integration was first pointed out in Granger (1981). A theorem showing that co-integrated series can be represented by error correction models was stated and proved in Granger (1983) and therefore is called the Granger

Representation Theorem. Analysis of related but more complex cases is covered by Johansen (1985) and Yoo (1985).

Section 3.3 Estimating Cointegration System

Besides maximum likelihood estimation procedure, with different representations for cointegrated systems, other estimation procedures have been proposed. The most convenient methods use the error correction form, especially when we can assume there is no moving average term. Two of these methods are decribed below.

The presentation of error correction model:

$$\Delta\mathbf{x}_t = \boldsymbol{\delta}\mathbf{x}_{t-1} + \boldsymbol{\Phi}^*{}_1\Delta\mathbf{x}_{t-1} + \cdots + \boldsymbol{\Phi}^*{}_{p-1}\Delta\mathbf{x}_{t-p+1} + \mathbf{a}_t \qquad (3.3.1)$$

naturally leads to a method of regression to get the estimate of $\boldsymbol{\delta}$. Johansen (1994) introduced a three-stage regression procedure:

1. Regress $\Delta\mathbf{x}_t$ on $\Delta\mathbf{x}_{t-1},\cdots,\Delta\mathbf{x}_{t-p+1}$ to obtain residual matrix $\mathbf{e}_{1t}$.
2. Regress $\mathbf{x}_{t-1}$ on $\Delta\mathbf{x}_{t-1},\cdots,\Delta\mathbf{x}_{t-p+1}$ to obtain residual matrix $\mathbf{e}_{2,t-1}$.
3. Regress $\mathbf{e}_{1t}$ on $\mathbf{e}_{2,t-1}$ to obtain the estimate of matrix $\boldsymbol{\delta}$.
4. Then estimate model (3.3.1) with $\boldsymbol{\delta}$ fixed at the estimated value obtained in step3 to get estimates of the $\boldsymbol{\Phi}^*{}_j$ s.

Notice $\boldsymbol{\delta} = -\boldsymbol{\gamma}\boldsymbol{\alpha}'$, so by examining the rank of $\boldsymbol{\delta}$ through its eigenvalues, we can also estimate and test the rank of cointegrating space.

Engle and Granger (1987) suggested another estimation method which is called two-step estimator. In the first step cointegration vector is estimated. And then the estimated cointegration vector is used in the error correction form to estimate the dynamics of the process. These two steps both require only ordinary least squares and the result is consistent for all the parameters. This estimating procedure is convenient in the sense that the dynamics do not need to be specified until the error correction structure has been estimated, and it also provides some test statistics useful for testing for cointegration.

If the p-dimensional vector process $\mathbf{x}'_t = [x_{1,t}, x_{2,t},\cdots,x_{p,t}]$ is $CI(1,1)$ with single cointegrating vector, there is a nonzero px1 vector $\boldsymbol{\alpha}' = [c_1, c_2,\cdots,c_p]$ such that $\boldsymbol{\alpha}'\mathbf{x}_t$ is stationary. Without loss of generality, say $c_1 \neq 0$. Then $(1/c_1)\boldsymbol{\alpha}'$ is also a cointegration vector, with the first element 1. Thus it is natural to consider the following regression model with $x_{1,t}$ as the dependent variable and $x_{2,t},\cdots,x_{p,t}$ the predictors:

$$x_{1,t} = \phi_1 x_{2,t} + \cdots + \phi_{p-1} x_{p,t} + \varepsilon_t .$$

This regression is called the cointegrating regression. It attempts to fit the long run equilibrium relationship without worrying about the dynamics. It provides an estimate of the elements of the cointegrating vector. Such a regression has been called a spurious regression by Granger and Newbold (1974) since the standard errors of the estimated regression coefficients are incorrect. So here we only seek coefficient estimates to use in the second stage of estimation and for tests of the equilibrium relationship. Further discussions about more general cases of more than one cointegration vectors can be found in Engle and Granger (1987).

The estimated cointegrating vector obtained by regression method provides a good approximation to the true cointegrating vector because it seeks vector with minimal residual variance. Asymptotically all linear combinations of $\mathbf{x}_t$ will have infinite variance except those which are cointegrating vectors. A point need to be made is that we estimate the cointegrating vector by normalizing the first element to be unity. However, we can normalize any nonzero element c_i and regress $x_{i,t}$ on other variables in estimating the regression coefficients. The results are invariant of the choice of $x_{i,t}$ as the dependent variable in the regression for most of the cases, but could be inconsistent sometimes. This is a weakness of this approach. But due to its simplicity, it is still commonly used.

In the second step, the remainder of the parameters of the cointegrated system are estimated by regressing the difference vector series on its lagged series and the equilibrium error term $\mathbf{z}_{t-1}$ with $\boldsymbol{\alpha}$ fixed at the estimated value in the computation of $\mathbf{z}_{t-1} = \alpha ' \mathbf{x}_{t-1}$. This simplifies the estimation procedure by imposing cross-equation restrictions and the dynamics of the system does not have to be specified in order to estimate α .

Surprisingly, the two-step estimator has excellent properties. As stated in the theorem below, it is as efficient as the maximum likelihood estimator based on the known value of α . Under some regular conditions the estimator is asymptotically normal. This theorem is first stated and proved by Engle and Granger (1987).

Theorem The two step estimator of a single equation of an error correction system, obtained by taking $\hat{\alpha}$ from the cointegrating regression as the true value, will have the same limiting distribution as the maximum likelihood estimator using the true value of α. Least squares standard errors will be consistent estimates of the true standard errors.

A simple example will illustrate this estimation procedure. Suppose two series are generated according to the following model:

$$\begin{aligned} x_{1t} + \beta x_{2t} &= u_{1t}, \quad u_{1t} = u_{1t-1} + \varepsilon_{1t}, \\ x_{1t} + \alpha x_{2t} &= u_{2t}, \quad u_{2t} = \rho u_{2t-1} + \varepsilon_{2t}, \quad |\rho| < 1 \end{aligned} \quad (3.3.2)$$

where ε_{1t} and ε_{2t} are white noise processes. In the usual sense, α and β are unidentifiable since there are no exogenous variables and the errors are contemporaneously correlated. Also, notice that $u_{2t} \sim I(0)$ and $u_{1t} \sim I(1)$. By simply rearranging terms, x_{1t} and x_{2t} can be expressed as linear combinations of u_{1t} and u_{2t}, so they are both $I(1)$. The second equation suggests that $x_{1t} + \alpha x_{2t}$ is a stationary series. Thus x_{1t} and x_{2t} are $CI(1,1)$. We will estimate the parameters by the two step approach.

First step, a linear least squares regression of x_{1t} on x_{2t} provides a good estimate of α. This is called cointegrating regression. All linear combination of x_{1t} and x_{2t} except $x_{1t} + \alpha x_{2t}$ defined in the model will have infinite variance. Therefore, it makes sense that regression of x_{1t} on x_{2t} by method of least square will give good estimate of α. For series generated by model (3.3.2), the reverse regression of x_{2t} on x_{1t} has the same property and will give a consistent estimate of $1/\alpha$.

Once the parameter α has been estimated, the others can be estimated in many ways conditional on the estimate of α. Let $\delta = (1-\rho)/(\alpha - \beta)$, then the generating model can be written in the autoregressive representation as

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \beta\delta & \alpha\beta\delta \\ -\delta & -\alpha\delta \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad (3.3.3)$$

where the η s are linear combinations of the ε s and thus are white noise themselves. Let $z_t = x_{1t} + \alpha x_{2t}$. Then model (3.3.3) can be written in the error correction representation form:

$$\begin{pmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{pmatrix} = \begin{pmatrix} \beta\delta \\ -\delta \end{pmatrix} z_{t-1} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}.$$

There are 3 unknown parameters in the original model (3.3.2). Now the error correction form has only 2 unknown parameters left. Once α is estimated in the first step, there is no constraints in the error correction model, thus we can get estimators for the dynamics system by simple regression or MLE. Notice that when $\rho \to 1$, the series are no longer cointegrated, but correlated random walks.

CHAPTER 4 - Test of Cointegration

It is usually of interest to test whether a set of variables are cointegrated. This may be desirable because of practical inquiries such as whether a system is in some form of equilibrium in the long run, and whether it is sensible to identify cointegration before estimating a multivariate dynamic model.

Unfortunately, the setup of cointegration system renders direct application of likelihood base test impossible. The testing of cointegration is closely related to tests for unit roots in observed series as formulated by Fuller (1976) and Dickey and Fuller (1979, 1981). It is also related to the problem of testing when some parameters are unidentified under the null hypothesis as discussed by Davies (1977) and Watson and Engle (1982).

In testing for cointegration in $\mathbf{x}_t$, sometimes we are particularly interested in a matrix or vector $\boldsymbol{\alpha}'$ based on some theoretical consideration. Then we can simply formulate the null hypothesis to test whether the process $\mathbf{z}_t = \boldsymbol{\alpha}'\mathbf{x}_t$ contains a unit root so that Dickey and Fuller test or Augmented Dickey and Fuller test is applicable. The distribution in this case is already nonstandard and was obtained through a simulation by Dickey (1976). We will conclude that $\mathbf{x}_t$ is cointegrated if the null hypothesis of unit roots is rejected. However, when $\boldsymbol{\alpha}'$ is unknown and estimated from the data, the Dickey-Fuller test tends to reject the null hypothesis too often. The reasons are that when the series is not cointegrated, $\boldsymbol{\alpha}'$ is not identifiable and that the variation of the estimated $\boldsymbol{\alpha}'$ is not accounted for.

Section 4.1 Seven Types of Tests

Suppose the true system is a bivariate linear vector autoregression with Gaussian errors where each of the series is individually $I(1)$ denoted by (x_t, y_t), Engle and Granger (1987) introduced seven types of tests. Each type is useful under some assumptions.

1. CRDW. After running the cointegrating regression, the Durbin Watson test is carried out to see if the residuals appear stationary. If they are nonstationary, the Durbin Watson statistic will approach zero and thus the test rejects non-cointegration null hypothesis if DW is too big. This was first proposed by Bhargava (1984) for the case when null and alternative are first order models.
2. DF. This tests the residuals from the cointegrating regression by running an auxiliary regression as described by Dickey and Fuller. It also assumes that the model is of only first order.
3. ADF. The augmented Dickey Fuller test allows for more lagged terms in the regression and is appropriate to use when higher order lags are needed.
4. RVAR. The restricted vector autoregression test is closely related to the two step estimator. Based on the estimate of the cointegrating vector from the cointegration regression, the error correction representation is estimated. Then whether the error correction term is significant is tested. First order system is assumed in this case.
5. ARVAR. The augmented RVAR test comes with the same idea as RVAR but allows higher order system.
6. UVAR. The unrestricted VAR test is based on a vector autoregression in the levels which is not restricted to satisfy the cointegration constraints.The test is simply whether the lagged levels would appear at all, or whether the model can be expressed entirely in changes. This test assumes first order model.
7. AUVAR. This is a higher order version of UVAR test.

The test statistics of the above seven types of tests are stated in Table 4.1. They are all computable by least squares. The critical values were estimated for each statistics by simulation using 10,000 replications by Engle and Granger (1987) under the null hypothesis of two independent $I(1)$ series. Using these critical values, the power of the test statistics were computed by simulations under various alternatives.

In the more complicated but realistic case that the system is of infinite order but can be approximated by a p th order autoregression, the statistics will only be asymptotically similar. Therefore, tests 3, 5 and 7 are asymptotically similar if the p th order model is true, whereas tests 1, 2, 4, and 6 are not asymptotically similar since these tests omit the lagged terms in regression. For this reason, Engle and Granger (1987) suggested one should not use the latter tests unless first order assumption is appropriate. Whether it is preferable to use a data base selection of p for these testing procedures needs further investigation. Furthermore, by comparing the critical values and powers for the seven tests under first order system and fourth order system assumptions, they decided that CRDW test is the best in power for first order case but too sensitive to changes of parameters in the null hypothesis. However, due to its simplicity,

Table 4.1 The Test Statistics: Reject for large values

1. The Cointegrating Regression Durbin Watson:. $y_t = \alpha x_t + c + u_t$.

$\xi_1 = DW.$ Under null hypothesis $DW = 0$

2. Dicky Fuller Regression: $\Delta u_t = -\phi u_{t-1} + \varepsilon_t$.

$\xi_2 = \tau_\phi$: the t statistic for ϕ .

3. Augmented DF Regression: $\Delta u_t = -\phi u_{t-1} + \Delta u_{t-1} + \cdots + \Delta u_{t-p} + \varepsilon_t$.

$\xi_3 = \tau_\phi$.

4. Restricted VAR: $\Delta y_t = \beta_1 u_{t-1} + \varepsilon_{1t}$, $\Delta x_t = \beta_2 u_{t-1} + \gamma \Delta y_t + \varepsilon_{2t}$.

$\xi_4 = \tau^2{}_{\beta 1} + \tau^2{}_{\beta 2}$.

5. Augmented Restricted VAR: Same as (4) but with p lags of Δy_t and Δx_t in each equation.

$\xi_5 = \tau^2{}_{\beta 1} + \tau^2{}_{\beta 2}$

6. Unrestricted VAR: $\Delta y_t = \beta_1 y_{t-1} + \beta_2 x_{t-1} + c_1 + \varepsilon_{1t}$, $\Delta x_t = \beta_3 y_{t-1} + \beta_4 x_{t-1} + \gamma \Delta y_t + c_2 + \varepsilon_{2t}$.

$\xi_6 = 2[F_1 + F_2]$

where F_1 is the F statistic for testing β_1 and β_2 both equal to zero in the first equation;

and F_2 is the F statistic for testing β_3 and β_4 both equal to zero in the second.

7. Augmented Unrestricted VAR: Same as (6) but with p lags of Δx_t and Δy_t in each equation.

$\xi_7 = 2[F_1 + F_2]$

CRDW is frequently used as a quick approximate result. Considering that realistically, one could not know which critical value to use, the ADF test with relative high power and quite consistent critical values for both first order and fourth order cases was recommended by Engle and Granger (1987) and has been widely used for testing cointegration. However, its power is slightly lower than DF test when first order can be assumed to be true. The critical values obtained by Engle and Granger for CRDW, DF and ADF test statistics are listed in Table 4.2.

The critical values listed here have only been estimated by simulation for the bivariate case for one sample size and from two specific models under null hypothesis. More general cases are remained to be discussed. Nevertheless, the critical values given in Table 4.2 have been used widely as a rough guide in applied studies.

Table 4.2 Critical Values: Reject for large values

First Order Model:

Δx_t , Δy_t independent standard normal, 100 observations, 10,000 replications, p=4

		Critical Values		
Statistics	Type of Test	1%	5%	10%
1	CRDW	0.511	0.386	0.322
2	DF	4.07	3.37	3.03
3	ADF	3.77	3.17	2.84

Higher Order Model:

$\Delta y_t = 0.8 \Delta y_{t-4} + \varepsilon_t$, $\Delta x_t = 0.8 \Delta x_{t-4} + \eta_t$

ε_t , η_t independent standard normal, 100 observations, 10,000 replications, p=4

		Critical Values		
Statistics	Type of Test	1%	5%	10%
1	CRDW	0.455	0.282	0.209
2	DF	3.90	3.05	2.71
3	ADF	3.73	3.17	2.91

Section 4.2 Critical Values Simulated for other sample sizes

To discuss the effect of sample sizes on the simulation results, critical values of DF and ADF tests are obtained by simulation under the same null hypotheses as in Table 4.2 but with various sample sizes. The independent series were generated according to the models under null hypothesis, then test statistics were calculated as stated in Table 4.1. The procedure was replicated for 10,000 times and the $(1-\alpha)$th percentiles were recorded as the critical values. Results is shown in Table 4.3.

Table 4.3 Critical Values for different sample sizes

First Order Model:

Δx_t , Δy_t independent standard normal, 100 observations, 10,000 replications, p=4

Type of Tests	DF			ADF		
	Critical Values			Critical Values		
Sample Size	1%	5%	10%	1%	5%	10%
30	4.37	3.55	3.19	3.77	3.04	2.71
50	4.13	3.50	3.15	3.74	3.14	2.83
70	4.05	3.43	3.11	3.82	3.22	2.90
80	4.03	3.43	3.11	3.89	3.26	2.95
90	4.01	3.38	3.07	3.80	3.23	2.94
100	3.99	3.39	3.08	3.80	3.25	2.93

Higher Order Model:

$\Delta y_t = 0.8\Delta y_{t-4} + \varepsilon_t$, $\Delta x_t = 0.8\Delta x_{t-4} + \eta_t$

ε_t , η_t independent standard normal, 100 observations, 10,000 replications, p=4

Type of Tests	DF			ADF		
	Critical Values			Critical Values		
Sample Size	1%	5%	10%	1%	5%	10%
30	5.66	4.67	4.14	3.90	3.21	2.86
50	5.39	4.27	3.70	3.87	3.26	2.95
70	4.83	3.73	3.25	3.93	3.31	3.00
80	4.69	3.68	3.19	3.87	3.30	3.00
90	4.47	3.51	3.06	3.88	3.33	3.01
100	4.34	3.48	3.02	3.89	3.31	3.02

It can be seen from Table 4.3 that the estimated critical values stabilized towards a limiting value as n approaches 100 and one should be cautious when use the critical values for n=100 if the sample size is less than 90.

Surprisingly, the critical values for sample size 100 are slightly larger than the critical values provided by Engle and Granger (1987). Thus, based on critical values in Table 4.3, it would be harder to reject the null hypothesis and thus conclude cointegration less often than based on critical values in Table 4.2. Consequently, the power of the test is lower. Since the algorithm of simulation has been checked carefully, it is possible that Engle and Granger (1987) used slightly different test statistic from the regression.

Section 4.3 Critical Values Simulated for other models

In reality, we are not likely to know beforehand which model is appropriate and thus which critical value to use. Hence, an ideal test statistic should be consistent under various kinds of null hypothesis. As the most widely used cointegration test DF and ADF, it is desirable to know their behavior under models with different coefficients, different lags and different forms. In this section, the critical values of ADF and DF tests obtained by simulation under various null hypotheses are tabulated in Table 4.4. Here five models were used as null hypotheses: the first one has only lag 4 term with coefficient 0.8 (this is the one used in Table 4.2); the second model includes all lagged terms with orders lower or equal to 4; the third one contains only lag 5 term; the fourth model is the same as model 3 with different coefficient; and the last one is an invertible moving average model.

It can be seen that the critical values of DF test vary dramatically with different models. Since DF test does not include any lagged term in the test regression, so not surprisingly it is sensitive to changes in lags. For ADF test, adding lower order lagged terms into the model doesn't affect the critical values very much. But with higher order lag, the critical values decrease, so that it would be easier to reject null hypotheses and detect cointegration. Consequently, if we were to use the critical values from Table 4.2 to detect cointegrated system with order higher than four, we might fail to reject the null hypothesis sometimes. Fortunately, autoregressive time series with more than 4 lags are not common.

Table 4.4 Critical Values for different models

100 observations, 10,000 replications, p=4 , ε_t , η_t independent standard normal						
Type of Tests	DF			ADF		
	Critical Values			Critical Values		
Models under Null Hypothesis	1%	5%	10%	1%	5%	10%
$\Delta y_t = 0.8\Delta y_{t-4} + \varepsilon_t$, $\Delta x_t = 0.8\Delta x_{t-4} + \eta_t$	4.34	3.48	3.02	3.89	3.31	3.02
$(1-0.8B)^4\Delta y_t = \varepsilon_t$, $(1-0.8B)^4\Delta x_{t-4} = \eta_t$	7.81	5.28	4.02	3.83	3.20	2.88
$\Delta y_t = 0.8\Delta y_{t-5} + \varepsilon_t$, $\Delta x_t = 0.8\Delta x_{t-5} + \eta_t$	4.58	3.62	3.16	3.78	2.78	2.31
$\Delta y_t = 0.7\Delta y_{t-5} + \varepsilon_t$, $\Delta x_t = 0.7\Delta x_{t-5} + \eta_t$	4.24	3.46	3.03	3.54	2.67	2.22
$\Delta y_t = (1-0.5B)(1-0.9B)\varepsilon_t$, $\Delta x_t = (1-0.8B)(1-0.9B)\eta_t$	18.08	16.83	16.24	6.56	5.89	5.55

It is worth noticing that when the 2 independent differenced series under null hypothesis are invertible moving average process instead of autoregression process, the critical values are much larger than those for model one. Hence, when the true system involves moving average term, tests based on critical values provided by Engle and Granger (1987) would reject the null too often with too many false positives.

This discussion is still based on the bivariate case and leaves many questions unanswered. Critical values for more variables and sample sizes were calculated by Engle and Yoo (1986) using the same general approach. Research on the limiting distribution theory by Phillips (1985) and Phillips and Durlauf (1985) might lead to alternative approach with better performance. We should be cautious if the structure of practical time series is not autoregressive or the test statistics is on the edge of critical values when applying test of cointegration.

CHAPTER 5 - Applications to Finance Data

Nowadays, due to economic globalization, decisions and activities taking place in one part of the world have significant impact for people and communities elsewhere in the world. This close relationship among different parts of the world can be seen in various kinds of economic and financial indexes and criteria.

To investigate the currency relationship between United States and Asia, the monthly log exchange rates of US Dollar (USD) vs British Pound (GBP) and the log exchange rate of Chinese Yuan (CNY) vs British Pound (GBP) from 1995 Jan to 2007 Dec were obtained. Initial plots of the two time series Figure 5.1 and Figure 5.2 show that they are both nonstationary but share similar trend over time as seen in Figure 5.3 (dotted line represents USD, and solid line represents CNY).

Figure 5.1 Log USDvsGBP exchange rate (1995-2007)

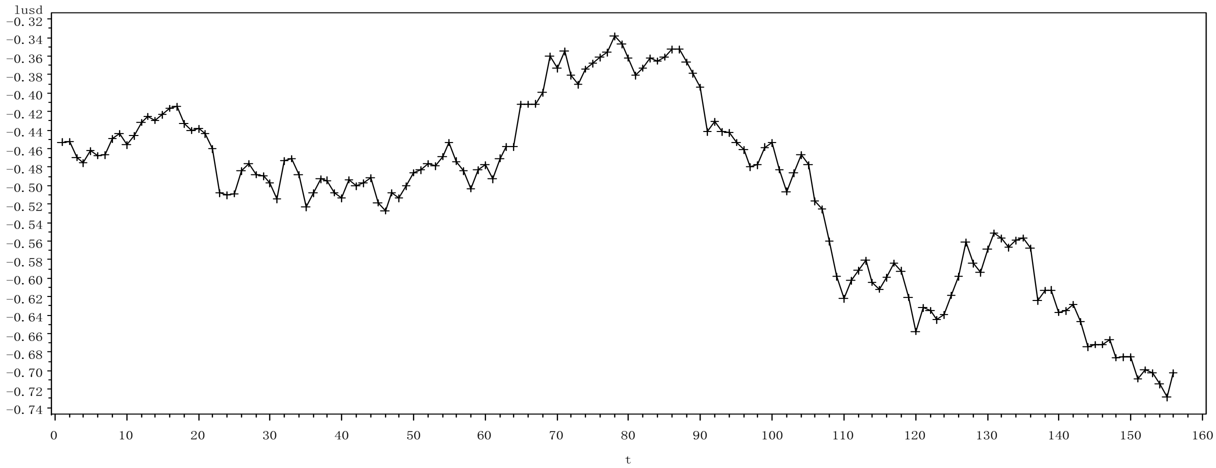
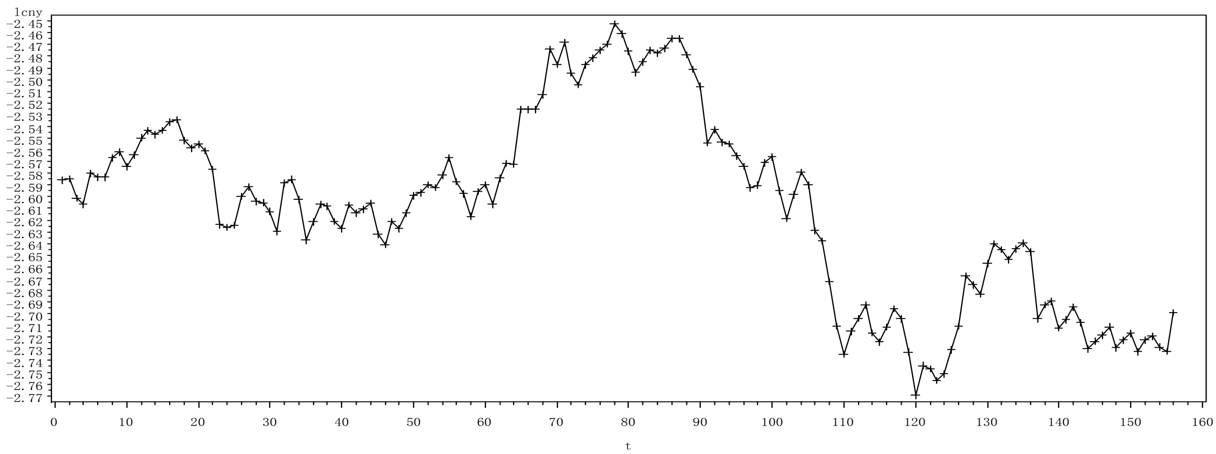
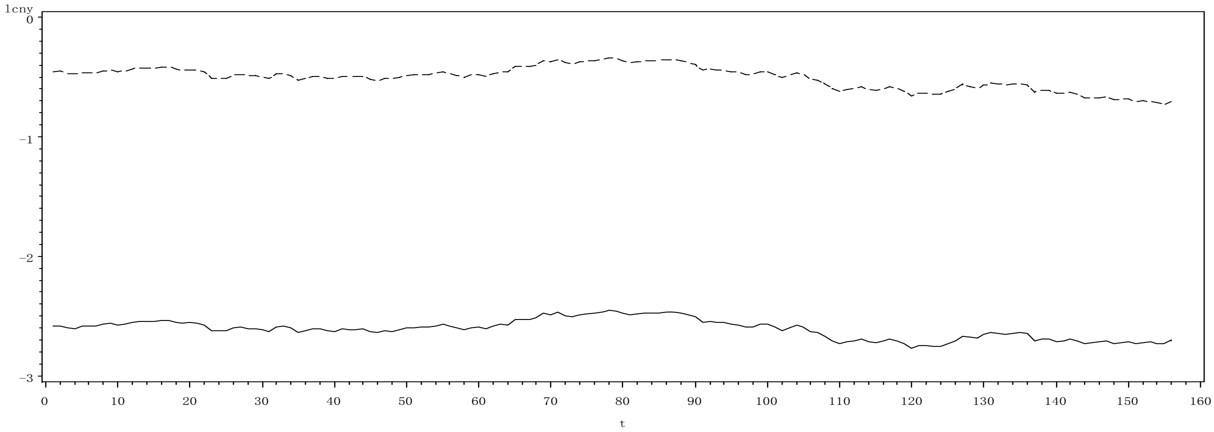


Figure 5.2 Log CNYvsGBP exchange rate (1995-2007)



Then it was checked that both series are $I(1)$. ADF test was run for log USDvsGBP exchange rate (denoted by LUSD) with lag 3. It gave a t-statistic -0.09 which suggests the existence of unit root. Running the same test for the first difference of the series with lag 2 yielded a t-statistic -7.47 indicating that first difference is stationary. For log CNYvsGBP

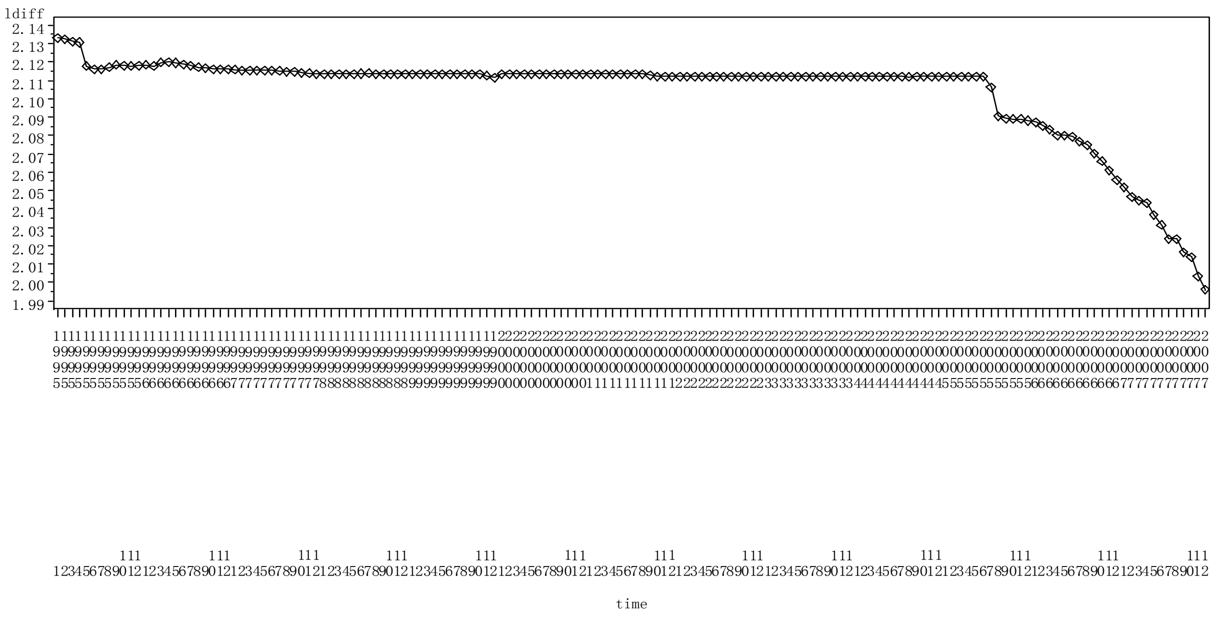
Figure 5.3 Log CNYvsGBP and USDvsGBP (1995-2007)



exchange rate (denoted by LCNY), same tests were used and two t-statistics were -1.06 and -7.35 respectively. Hence both series are $I(1)$.

It is of interest to know if the ratio of USD and CNY remains stationary over time. Then a test for whether log ratio=LUSD-LCNY (denoted as LDIFF) is stationary or not could be conducted. In this case, the cointegration vector for testing is known as (1, -1). Thus an ADF or DF test on the series of difference between log USDvsGBP exchange rate and log CNYvsGBP exchange rate would be sufficient. Surprisingly, ADF test with first lag gave a t-statistic 5.20

Figure 5.4 The Plot of Residuals of Cointegrating Regression (LDIFF)



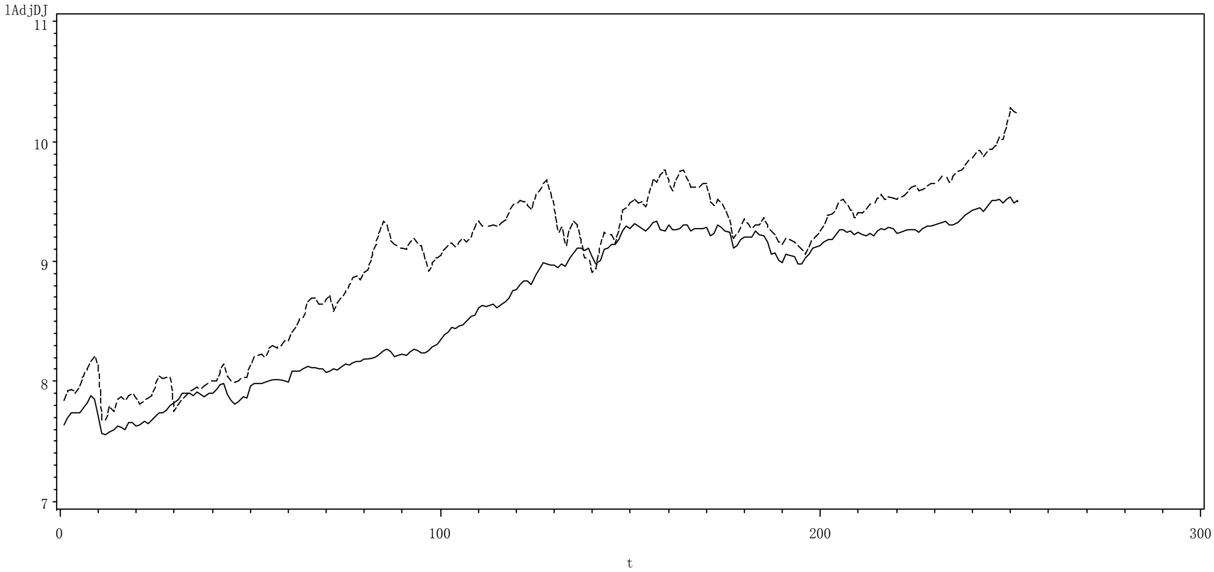
indicating that one cannot reject the null hypothesis, thus the two series were not cointegrated by vector (1, -1). The LDIFF series is plotted to show its behavior. It seems that there was a gradually drop in the log ratio between USD and CNY from 2005's June till the end of 2007 as shown in Figure 5.4. Therefore, it is possible that equilibrium exists but has been violated by an event happened around 2005 June.

By removing the observations since 2005's June till now, the remain two series are still $I(1)$. (ADF tests with 2 lags for LUSD and LCNY yielded t statistics -0.67 and -0.69 indicating that both series are nonstationary. And DF tests for the first differenced series gave t-statistics -9.57 and -9.66 suggesting that they are stationary after once difference.) LDIFF was tested again for stationarity by ADF test with first lag. The test statistic turned out to be -6.17 suggesting that one should reject the unit root null hypothesis and conclude stationarity. Now LDIFF=LUSD-LCNY is stationary, we can conclude that LUSD and LCNY are cointegrated with cointegration vector (1, -1) until 2005 June. It is found that before 2005, the value of China's Currency, the Yuan has been linked to US Dollar through government adjustment. However, in the June of 2005, China's political leadership actively mentioned the thought of breaking such a link and instead, tying Yuan's value to a group of currencies as Euro, Yen etc. Since then, the value of CNY has been slowly but steadily going up causing the log ratio of USD over CNY decreased and broke the equilibrium of the past ten years. Although losing the equilibrium with USD, CNY is very likely to be cointegrated with average values of the dollar, yen, euro and possibly other currencies like the British pound.

Other than currency exchange rates, the performance of stock markets also presents certain relationships in economics and finance between different districts of the world. Here, the monthly average of Adjusted Close quotes (The adjusted close adjusts for dividends and stock splits for the stock and will be a different number than the close.) of Dow Jones Industrial Average and the Hong Kong HSI-HANG SENG from 1987 Jan to 2007 Dec were put together to check the underlying relationship between the two stock indexes. Although they are both going up, there is no clear common trend can be detected from the graph. Hence, they might not be cointegrated. Tests were conducted to see if cointegration can be detected.

First, the two series were taken log to stabilize the variances and plotted together with each other. The solid line presents the behavior of log Adjusted close quotes of Dow Jones Industrial Average, while the dotted line presents that of HSI-HANG SENG.

Figure 5.5 Plots of Log Adjusted Close quotes of Dow Jones and His-hang Seng



They are denoted by LAdjDJ and LAdjHS. Then they were checked by ADF tests for stationarity. The ADF test with lag 2 of the original series gave t-statistics -0.71 for LAdjDJ and -0.76 for LAdjHS suggesting neither of them is stationary; after the first order difference, they were tested again by ADF tests with lag 4, which gave a t-statistic -7.41 for LAdjDJ and -8.36 for LAdjHS indicating that after first difference, both series turned out to be stationary. Thus, the two time series of interest are both $I(1)$. Cointegrating regression was then run and DW turned out to be 0.0413 which is not even close to the critical value listed in Table 4.2. A regression of the differenced residual series on one lagged residual and 4 lags of the differenced terms was then run and the results of both regressions are shown in Table5.1. The ADF test statistic is -1.51. Based on the critical values listed in Table 4.2, we cannot reject the non-cointegration null hypothesis and thus failed to detect cointegration between Log AdjDJ and Log AdjHS. Since all the lagged terms appear to be significiant at least under 0.10 level, using DF test to seek for higher power is not appropriate in this case. Further, when the cointegrating regression was reversed, by regressing Log AdjHS on Log AdjDJ, similar results were obtained and no cointegration was identified. In conclusion, the log Adjusted Close quotes for Dow Jones Industrial Average and that for Hong Kong HSI-HANG SENG are not cointegrated. It is not so

Table 5.1 Regression of LAdjcloseDJ on LAdjcloseHS

Dependent Variables	Independent Variables, estimates and t-stats						
	c	LAdjHS	Res(-1)	Δ Res(-1)	Δ Res(-2)	Δ Res(-3)	Δ Res(-4)
LAdjDJ	0.783 (3.73)	0.873 (37.63)					
Δ Res			-0.0196 (-1.51)	0.1697 (2.66)	-0.1189 (-1.85)	0.1158 (1.80)	-0.1026 (-1.59)

surprisingly in the sense that instead of close relationship between just two stock markets, mutually impacts and constraints are expected as economics globalized. Hence, we might expect that if more variables such as the quotes for Shang Hai Stock Market or Nasdaq were available, then they might be cointegrated. Furthermore, it is clear from the plots in Figure 5.5 that the Hong Kong stock market is more volatile than the US market and the investment patterns of the two eareas are different due to different types of investors. This may also explain why the two series are not cointegrated.

References And Bibliography

Bhargava, Alok (1984): “On the Theory of Testing For Unit Roots in Observed Time Series,” manuscript, ICERD, London School of Economics.

Box, G. E. P. and Tiao, G. C. (1977): “A Canonical Analysis of Multiple Time Series,” Biometrika (1977), 64, 2, 355-65.

Davidson, J. E. H., David F. Hendry, Frank Srba, and Steven Yeo (1978): “Econometric Modelling of the Aggregate Time-series Relationship Between Consumer’s Expenditure and Income in the United Kingdom,” Economic Journal, 88, 661-692.

Davies, R. R. (1977): “Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative,” Biometrika, 64, 247-254.

Dickey, David A. (1976): “Estimation and Hypothesis Testing for Nonstationary Time Series,” PhD. Thesis, Iowa State University, Ames.

Dickey, David A, and Wayne A. Fuller (1979): “Distribution of the Estimators for Autoregressive Time Series With a Unit Root,” Journal of the American Statistical Assoc., 74, 427-431.

--------- (1981): “The Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root,” Econometrica, 49, 1057-1072.

Engle, Robert F., and Byung SamYoo (1986): “Forcasting and Testing in Co-integrated Systems,” U.C.S.D. Discussion Paper.

Engle, Robert F., and Granger C. W. J. (1987): “Co-integration and Error Correction: Representation, Estimation, and Testing,” Econometrica, Vol.55, No.2 (Mar., 1987), 251-276.

Feller, William (1968): *An Introduction to Probability Theory and Its Applications*, Volume I. New York: John Wiley.

Fuller, Wayne A. (1976): *Introduction to Statistical Time Series*. New York: John Wiley.

Granger, C. W. J. (1981): “Some Properties of Time Series Data and Their Use in Econometric Model Specification,” Journal of Econometrics, 121-130.

----------(1983): “Co-Integrated Variables and Error-Correcting Models,” unpublished UCSD Discussion Paper 83-13.

Granger, C. W. J., and P. Newbold (1977): Forecasting Economic Time Series. New York: Academic Press.

--------(1974): “Spurious Regressions in Econometrics,” Journal of Econometrics, 26, 1045-1066.

Granger, C. W. J., and A. A. Weiss (1983): “Time Series Analysis of Error-Correcting Models,” in *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press, 255-278.

Javier López-de-Lacalle and Ignacio Díaz-Emparanza (2005). uroot: Unit Root Tests and Graphics for Seasonal Time Series. R package version 1.4.

Johansen, Soren (1985): “The Mathematical Structure of Error Correction Models,” manuscript, University of Copenhagen.

---------(1994): “Estimating Systems of Trending Variables,” Econometric Reviews, Taylor and Francis Journals, vol. 13(3), pages 351-386.

Phillips, P. C. B. (1985): “Time Series Regression with Unit Roots,” Cowles Foundation Discussion Paper No. 740, Yale University.

Phillips, P. C. B., and S. N. Durlauf (1985): “Multiple Time Series Regression with Integrated Processes,” Cowles Foundation Discussion Paper 768.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Tsay, Ruey S. (2002): “Analysis of Financial Time Series,” Wiley-Interscience, New York, USA. ISBN 0-471-41544-8.

Watson, Mark W., and Robert Engle (1985): “A Test for Regression Coefficient Stability with a Stationary AR(1) Alternative,” *The Review of Economics and Statistics*, Vol. 67, No. 2 (May, 1985), pp. 341-346.

Wei, William W.S.: “Time Series Analysis : Univariate and Multivariate Methods, 2nd” Addison Wesley (2005). ISBN 0-321-32216-9.

Yoo, Sam (1985): “Multi-co-integrated Time Series and Generalized Error Correction Models,” manuscript in preparation, U.C.S.D.

Appendix A - Codes for Simulation Study

The simulation was conducted by R.

 # ADF and DF test functions are available in the package "uroot";

library(uroot);

Generate Lagged I(1) series and get test statistics

 # Generate the series under null hypothesis;

$\Delta y_t = c\Delta y_{t-l} + \varepsilon_t$, $\Delta x_t = c\Delta x_{t-l} + \eta_t$;

 # Parameter in the model: order of the lag=l; coefficient=c, sample sizes=n;

```
lag4stat<-function(l,c,n)
{
        x<-numeric(); y<-numeric();
        deltax<-numeric();
        deltay<-numeric();
        e<-rnorm((n+500),0,1);
        u<-rnorm((n+500),0,1);
        deltax[1:l]<-e[1:l];
        deltay[1:l]<-u[1:l];
        for(i in (l+1):(n+500))
        {
                deltax[i]<-c*deltax[i-l]+e[i];
                deltay[i]<-c*deltay[i-l]+u[i];
                x[i]<-sum(deltax[1:i]);
                y[i]<-sum(deltay[1:i]);
```

```
        }
# Erase the first 500 observation;
        x<-x[501:(n+500)];
        y<-y[501:(n+500)];
# Regress y on x to get residual then get ADF and DF test statistics;
        lm<-lm(y~x);
        res<-resid(summary(lm));
        adfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
                selectlags=list(mode=c(1,2,3,4), Pmax=4));
        dfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
                selectlags=list(mode=c(0), Pmax=4));
        adf<-adfout@stat; df<-dfout@stat;
    return(c(adf[1,3],df[1,3]));
};

begin<-Sys.time();
# Assign number of replication and number of observation, i.e. sample sizes;
        rep<-10000;
        obs<-100;
        adf<-numeric(rep);
        df<-numeric(rep);
# Replicate the function 10,000 times and collect all ADF and DF stats;
        for(j in 1:rep)
        {
                adf[j]<-lag4stat(l=4,c=0.8,n=obs)[1];
                df[j]<-lag4stat(l=4,c=0.8,n=obs)[2];
        }
# Sort the stats and find the desired percentiles;;
        lag4adf<-c(sort(adf)[0.01*rep],sort(adf)[0.05*rep],sort(adf)[0.10*rep]);
        lag4df<-c(sort(df)[0.01*rep],sort(df)[0.05*rep],sort(df)[0.10*rep]);
        lag4adf;lag4df;
```

Generate No Lag I(1) series and get test statistics

 # Generate the series under null hypothesis;

Δy_t ***and,*** Δx_t ***are independent standard normal;***

 # Parameter: sample size=n;

```
nolagstat<-function(n)
{
	x<-numeric(); y<-numeric();
	deltax<-rnorm((n+500),0,1);
	deltay<-rnorm((n+500),0,1);
	for(i in 1:(n+500))
	{
		x[i]<-sum(deltax[1:i]);
		y[i]<-sum(deltay[1:i]);
	}
	x<-x[501:(n+500)];
	y<-y[501:(n+500)];
	lm<-lm(y~x);
	res<-resid(summary(lm));
	adfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
		selectlags=list(mode=c(1,2,3,4), Pmax=4));
	dfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
		selectlags=list(mode=c(0), Pmax=4));
	adf<-adfout@stat; df<-dfout@stat;
	return(c(adf[1,3],df[1,3]));
};
```

 # Replicate the procedure 10,000 times and collect all test stats;

```
	adf<-numeric(rep);
	df<-numeric(rep);
	for(j in 1:rep)
```

```
	{
		adf[j]<-nolagstat(n=obs)[1];
		df[j]<-nolagstat(n=obs)[2];
	}
```

 # Get desired percentiles as critical valus;

```
	nolagadf<-c(sort(adf)[0.01*rep],sort(adf)[0.05*rep],sort(adf)[0.10*rep]);
	nolagdf<-c(sort(df)[0.01*rep],sort(df)[0.05*rep],sort(df)[0.10*rep]);
	nolagadf;nolagdf;
```

 # To estimate the time needed for simulation;

```
end<-Sys.time();
end-begin;
```

Generate Lag 1,2,3,4 I(1) series and get test statistics

```
lag1234stat<-function(n)
{
	x<-numeric(); y<-numeric();
	deltax<-numeric();
	deltay<-numeric();
	e<-rnorm((n+500),0,1);
	u<-rnorm((n+500),0,1);
	deltax[1]<-e[1];
	deltay[1]<-u[1];
	deltax[2]<-3.2*deltax[1]+e[2];
	deltay[2]<-3.2*deltay[1]+u[2];
	deltax[3]<-3.2*deltax[2]-3.84*deltax[1]+e[3];
	deltay[3]<-3.2*deltay[2]-3.84*deltay[1]+u[3];
	deltax[4]<-3.2*deltax[3]-3.84*deltax[2]+2.048*deltax[1]+e[4];
	deltay[4]<-3.2*deltay[3]-3.84*deltay[2]+2.048*deltay[1]+u[4];
```

```
		for(i in 5:(n+500))
		{
			deltax[i]<-3.2*deltax[i-1]-3.84*deltax[i-2]+2.048*deltax[i-3]-
0.4096*deltax[i-4]+e[i];
			deltay[i]<-3.2*deltay[i-1]-3.84*deltay[i-2]+2.048*deltay[i-3]-
0.4096*deltay[i-4]+u[i];
			x[i]<-sum(deltax[1:i]);
			y[i]<-sum(deltay[1:i]);
		}
		x<-x[501:(n+500)];
		y<-y[501:(n+500)];
		lm<-lm(y~x);
		res<-resid(summary(lm));
		adfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
			selectlags=list(mode=c(1,2,3,4), Pmax=4));
		dfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
			selectlags=list(mode=c(0), Pmax=4));
		adf<-adfout@stat; df<-dfout@stat;
	  return(c(adf[1,3],df[1,3]));
	};

	##results;
	begin<-Sys.time();
		rep<-10000;
		obs<-100;
		adf<-numeric(rep);
		df<-numeric(rep);
		for(j in 1:rep)
		{
			adf[j]<-lag1234stat(n=obs)[1];
			df[j]<-lag1234stat(n=obs)[2];
```

```
        }
        lagladf<-c(sort(adf)[0.01*rep],sort(adf)[0.05*rep],sort(adf)[0.10*rep]);
        lagldf<-c(sort(df)[0.01*rep],sort(df)[0.05*rep],sort(df)[0.10*rep]);
        lagladf;lagldf;

end<-Sys.time();
end-begin;
```

Generate Invertible MA(2) I(1) series and get test statistics

```
library(uroot);
MA2stat<-function(n)
{
        x<-numeric(); y<-numeric();
        deltax<-numeric();
        deltay<-numeric();

        e<-rnorm((n+500),0,1);
        u<-rnorm((n+500),0,1);
        deltax[1]<-e[1];
        deltay[1]<-u[1];
        deltax[2]<-e[2]-1.7*e[1];
        deltay[2]<-u[2]-1.4*u[1];

        for(i in 3:(n+500))
        {
                deltax[i]<-e[i]-1.7*e[i-1]+0.72*e[i-2];
                deltay[i]<-u[i]-1.4*u[i-1]+0.45*u[i-2];
                x[i]<-sum(deltax[1:i]);
                y[i]<-sum(deltay[1:i]);
```

```
		}
		x<-x[501:(n+500)];
		y<-y[501:(n+500)];
		lm<-lm(y~x);
		res<-resid(summary(lm));
		adfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
			selectlags=list(mode=c(1,2,3,4), Pmax=4));
		dfout<-ADF.test(wts=ts(res), itsd=c(0,0,c(0)),regvar=0,
			selectlags=list(mode=c(0), Pmax=4));
		adf<-adfout@stat; df<-dfout@stat;
	return(c(adf[1,3],df[1,3]));
};

##results;
begin<-Sys.time();
		rep<-10000;
		obs<-100;
		adf<-numeric(rep);
		df<-numeric(rep);
		for(j in 1:rep)
		{
			adf[j]<-MA2stat(n=obs)[1];
			df[j]<-MA2stat(n=obs)[2];
		}
		adf<-c(sort(adf)[0.01*rep],sort(adf)[0.05*rep],sort(adf)[0.10*rep]);
		df<-c(sort(df)[0.01*rep],sort(df)[0.05*rep],sort(df)[0.10*rep]);
		adf;df;

end<-Sys.time();
end-begin;
```

Appendix B - Codes for Analysis of Finance Data

Examples were analyzed by SAS.

Currency Exchange Rate Data Analysis

Data can be obtained at *http://www.oanda.com/convert/fxhistory*

Arrange the data;

```
data sasuser.currency;
      merge meanusd meancny t;
      time=year||month;
      lusd=log(usdrate);
      lcny=log(cnyrate);
      ldiff=lusd-lcny;
      drop _TYPE_ _FREQ_;
run;
```

**** Plot the two series separately;***

```
proc gplot data=sasuser.currency;
      plot lcny*time lusd*time;
run;
```

**** Plot the two series together;***

```
goptions colors=(black);
      symbol1  i=join v=none l=1;
      symbol2  i=join v=none l=3;

proc gplot data=sasuser.currency;
      plot lcny*time lusd*time/overlay;
run;
```

**** Check if the two series are both I(1);***

****First check if the series are stationary;***

```
proc arima data=sasuser.currency;
      identify var=lusd stationarity=(ADF=(3));
      estimate p=4;
      identify var=lcny stationarity=(ADF=(3));
      estimate p=4;
run;
```

****Then check if the series after once difference are stationary;***

```
proc arima data=sasuser.currency;
      identify var=lusd(1) stationarity=(ADF=(3));
      estimate p=4;
      identify var=lcny(1) stationarity=(ADF=(3));
      estimate p=4;
run;
```

**** check if LUSD-LCNY is a stationary series;***

```
proc arima data=sasuser.currency;
      identify var=ldiff stationarity=(ADF=(1));
      estimate p=4;
run;
```

****Plot the LDIFF series;***

```
proc gplot data=res;
      symbol v=diamond i=join;
      plot ldiff*time;
run;
```

****Cutoff the irregular observations after 2005 June;***

```
data cutoff;
      set sasuser.currency;
      if t>126 then delete;
run;
```

****ADF tests for the series after cutoff to see if they are still I(1);***

```
proc arima data=cutoff;
      identify var=lusd stationarity=(ADF=(3));
      estimate p=4;
      identify var=lcny stationarity=(ADF=(3));
      estimate p=4;
run;

proc arima data=cutoff;
      identify var=lusd(1) stationarity=(ADF=(0));
      estimate p=4;
      identify var=lcny(1) stationarity=(ADF=(0));
      estimate p=4;
run;
```

****ADF tests for LDIFF after cutoff to see if it is stationary;***

```
proc arima data=cutoff;
      identify var=ldiff stationarity=(ADF=(1));
      estimate p=4;
run;
```

Stock Market Data Analysis

Data can be obtained at *http://finance.yahoo.com/*

Plot the two series DJ and NQ;

```
proc gplot data=sasuser.stock;
      plot lAdjDJ*t lAdjHS*t/overlay;
run;
```

**** Check if the two series are both I(1);***

First check if the series are stationary;

```
proc arima data=sasuser.stock;
      identify var=lAdjDJ stationarity=(ADF=(2));
      estimate p=4;
      identify var=lAdjHS stationarity=(ADF=(2));
      estimate p=4;
      identify var=lAdjNQ stationarity=(ADF=(2));
      estimate p=4;

run;
```

Then check if series after once difference are stationary;

```
proc arima data=sasuser.stock;
      identify var=lAdjDJ(1) stationarity=(ADF=(4));
      estimate p=4;
      identify var=lAdjHS(1) stationarity=(ADF=(4));
      estimate p=4;
      identify var=lAdjNQ(1) stationarity=(ADF=(1));
      estimate p=4;

run;
```

Cointegrating Regression is run to get DW and residuals;

```
proc autoreg data=sasuser.stock;
      model lAdjDJ=lAdjHS;
      output out=res residual=res;
run;
```

ADF test for the residual series from the cointegrating regression;

```
proc arima data=res;
      identify var=res stationarity=(ADF=(4));
      estimate p=4 noconstant;
run;
```

Get lagged terms from the data and Regression of the differenced residual series;

```
data DJHS;
```

```
      set res;
      ec=res;
      ec1=lag(ec);
      dec=ec-ec1;
      dec1=lag(dec);
      dec2=lag(dec1);
      dec3=lag(dec2);
      dec4=lag(dec3);
run;

proc autoreg data=djhs;
      model dec=ec1 dec1 dec2 dec3 dec4/noint;
run;
```

Plot the two series DJ and NQ;

```
goptions colors=(black);
      symbol1  i=join v=none l=1;
      symbol2  i=join v=none l=3;

proc gplot data=sasuser.stock;
      plot ladjDJ*t ladjNQ*t/overlay;
run;
```

Cointegrating Regression and the reverse regression;

```
proc autoreg data=sasuser.stock;
      model lAdjNQ=lAdjDJ;
      output out=res2 residual=res;
run;

proc autoreg data=sasuser.stock;
      model lAdjDJ=lAdjNQ;
      output out=res3 residual=res;
run;
```

ADF tests for the residuals to see if cointegration exists;

```
proc arima data=res2;
      identify var=res stationarity=(ADF=(4));
      estimate p=4;
run;

proc arima data=res3;
      identify var=res stationarity=(ADF=(4));
      estimate p=4;
run;
```

Get Lags and Differenced Series from the original Data;

```
data DJNQ;
      set res3;
      c=lAdjNQ;
      c1=lag(c);
```

```
      ec=res;
      ec1=lag(ec);
      dc=c-c1;
      dc1=lag(dc);
      dc2=lag(dc1);
      dc3=lag(dc2);
      dc4=lag(dc3);
      y=lAdjDJ;
      y1=lag(y);
      dy=y-y1;
      dy1=lag(dy);
      dy2=lag(dy1);
      dy3=lag(dy2);
      dy4=lag(dy3);
      dec=ec-ec1;
      dec1=lag(dec);
      dec2=lag(dec1);
      dec3=lag(dec2);
      dec4=lag(dec3);
run;
```

Regressions to estimate the final model;

```
proc autoreg data=djnq;
      model dec=ec1 dec1 dec2 dec3 dec4/noint;
run;

proc autoreg data=djnq;
      model dc=c1 y1 dc1 dc2 dc3 dc4 dy1 dy2 dy3 dy4;
run;

proc reg data=djnq;
      model dc=ec1 dc1 dy3 dy4/noint;
run;
```