DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

STATISTICAL INFERENCE OF COINTEGRATING VECTORS

by
Selim Orhun SUSAM

July, 2013
İZMİR

STATISTICAL INFERENCE OF COINTEGRATING VECTORS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for
The Degree of Master of Science in Statistics

by

Selim Orhun SUSAM

July, 2013

İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"STATISTICAL INFERENCE OF COINTEGRATING VECTORS"** completed by **SELİM ORHUN SUSAM** under supervision of **ASSOC. PROF. DR. ESİN FİRUZAN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

.....

Assoc. Prof. Dr. Esin FİRUZAN

Supervisor

.....

(Jury Member)

.....

(Jury Member)

Prof. Dr. Ayşe OKUR

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I am grateful to Assoc. Prof. Dr. Esin FİRUZAN, my supervisor, for their encouragement and insight throughout my research and for guiding me through the entire thesis process from start to finish. Has it not been for their faith in me, I would not have been able to finish the thesis.

I'm also grateful for the insights and efforts put forth by the examining committee; Prof. Dr. Vedat PAZARLIOĞLU and Assist. Prof. Dr. A. Fırat ÖZDEMİR.

I wish to give a heartfelt thanks to my father, Süleyman SUSAM, and my mother, Sibel SUSAM. They offered their endless support and constant prayers. This academic journey would not have been possible without their love, patience, and sacrifices along the way.

Selim Orhun SUSAM

STATISTICAL INFERENCE OF COINTEGRATING VECTORS

ABSTRACT

Cointegration analysis states that, in case the economic variable to be analyzed is not stationary, a linear combination of these series would be stationary. Put it differently, cointegration studies the linear combination of non-stationary variables. A simulation study is conducted in Chapter Four for the estimation of the coefficient matrix for the cointegrated vector autoregressive process. This study, in the last chapter, gives information about the performances of Johansen Trace and Maximum Eigenvalue tests, used for testing cointegration, depending on the size of the sample and the number of the variables in the system.

Keywords: Cointegration, least square method, maximum likelihood method, trace test, maximum eigenvalue test

EŞBÜTÜNLEŞME VEKTÖRÜ İÇİN İSTATİSTİKSEL ÇIKARSAMA

ÖZ

Eşbütünleşme analizi incelenen ekonomik değişkenin durağan olmaması durumunda, bu serilerden oluşturulan doğrusal bir birleşimin durağan olacağını ifade etmektedir. Yani başka bir ifadeyle eşbütünleşme durağan olmayan değişkenlerin doğrusal bir birleşimi ile ilgilenmektedir. Eşbütünleşik ikinci dereceden vektör otoresgresif sürecin katsayılar matrisinin tahmini için dördüncü bölümde simülasyon çalışması yapılmıştır. Bu çalışmanın son bölümünde eşbütünleşme testi için kullanılan Johansen trace ve maximum eigen value testlerinin örneklem büyüklüğüne ve sistemde yer alan değişken sayısına göre performansları hakkında bilgi vermektedir.

Anahtar sözcükler: Eşbütünleşme, maksimum olabilirlik yöntemi, en küçük kareler yötemi, iz testi, en büyük özdeğer testi

CONTENTS

Page

LIST OF FIGURES

Pages

LIST OF TABLES

Pages

CHAPTER ONE
INTRODUCTION

A time series is a sequence of the values a variable has for successive time units. Time series analysis is the statistical investigation of the data observed in time. It is possible to estimate the future values of the economic time series by making use of the values it had in the past. Time series can be discussed under two categories, stationary and non-stationary, with regard to the deviations from the mean value they exhibit.

In the econometric time series, the most important hypothesis for obtaining econometrically significant relations between the variables is the requirement for the time series to be analyzed to be stationary series. If the mean and variance of the time series do not change with regard to time, and if the covariance between two steps depends on the distance between these two steps rather than time stationarity is present. Therefore, stationarity concept has an important place in time series. On the contrary, economic time series show a tendency of increasing in time. This means that most of the economic time series are not stationary.

There are some consequences for using non-stationary time series variables in times series analyses. Spurious regression issue is one of the major issues and this would cause problems in hypothesis tests. In case of using stationary series these issues are solved to a great extent.

There are substantial differences between the stationary and non-stationary time series. A stationary time series shows an inclination of returning to the mean level in the long term. Covariance has a finite value and does not change in time. On the other hand, variance and mean of a non-stationary time series depend on time. Chapter two of this study summarizes the concept of stationarity in time series.

The domains in the time series field which have been focused recently are the studies that question the determination of the unit root, whether or not the time series are stationary, at which order the series are cointegrated and whether there is a cointegration relation between the series. If the series is not a stationary one, it should be made stationary using various methods. One of the main reasons of

making the series stationary is providing the hypothesis about the error terms. Therefore, a unit rooted series is tried to be made stationary by taking difference. However, taking difference method causes long term data loss in variables and it causes statistically erroneous results in the analysis of the series.

Cointegration analysis states that, in case the economic variables to be analyzed are not stationary, a linear combination of these series would be stationary. Put it differently, cointegration studies the linear combination of non-stationary variables.

Studies in which the cointegration analysis is used can be listed as follows: relations between expenses and revenues, relations between the long and short term interest rates, and the relations between the production and sales volume.

A simulation study is conducted in Chapter Four for the estimation of the coefficient matrix for the cointegrated vector autoregressive process. In this simulation, the asymptotic properties of $E(\hat{\rho}) - \rho$ ve $E(\hat{\alpha}) - \alpha$, for ρ and α values of the coefficient matrix A, are investigated for different conditions as per the method used for obtaining the unit root. For instance, the unit root in the process can be obtained in two ways depending on $\rho = 1$ and $\alpha < 1$ or $\alpha = 1$ and $\rho < 1$.

There are two separate hypothesis test methods frequently used for the determination of the existence of cointegration. These are the Engle-Granger and Johansen cointegration tests. Engle and Granger (1987) argues that if times with common trend are also integrated at the same order and the difference between the time series is stationary, then these series are cointegrated. Johansen (1988) method, on the other hand, is the multivariate generalization of the Engle-Granger method. While only one cointegration is found between variables in Engle-Granger method, more than one cointegration relations can be found in Johansen cointegration test. Also, while Engle-Granger test uses the least squares method for the estimation of the cointegration vector, Johansen test uses the Maximum Likelihood method.

This study, in the last chapter, gives information about the performances of Johansen Trace and Maximum Eigenvalue tests, used for testing cointegration, depending on the size of the sample and the number of the variables in the system. 2000 sample groups will be randomly generated, sample size being 30 and 400, using

a simulation method with varying f and g parameters. The presence of cointegration in the sample groups, generated with regard to the $\alpha = 0.05$ significance level which is mostly preferred in the literature, will be investigated with reference to the rejection ratios of H_0 hypothesis, and the power of the tests for each method will be questioned.

CHAPTER TWO
STATIONARY AND NON-STATIONARY TIME SERIES

2.1 Graphical Investigation of Stationary and Non-stationary Time Series

Stationary series have mean and variance that do not change in time. Such series would present a constant oscillation and move around its own mean. For nonstationary series, on the other hand, the variance of the series becomes a function of time. All statistical tests give correct results under the assumption that the series is a stationary one.

For stating the series x_t , defined for t=1,2,…,T times, as stationary, the three conditions given below should be met:

a) Constant mean, $E(X_t) = E(X_s) = \mu \ \ ,\forall \ t \neq s$

b) Constant and finite variance, $Var(X_t) = Var(X_s) = \sigma^2 < \infty \ ,\forall \ t \neq s$

c) $Cov(X_t, X_s) = \Gamma_{t-s}$ ve $Cov(X_{t+i}, X_{s+i}) = \Gamma_{t-s}$, $\forall \ i$

The series which do not have a constant oscillation around a constantmean and which do not satisfy the three conditions above are called the nonstationary time series. The stationarity of the AR (1) model given in Equation 2.1 depends on the a_1 coefficient.

$$x_t = a_1 x_{t-1} + \varepsilon_t \tag{2.1}$$

If $|a_1| < 1$, then the process is called stationary; and if $a_1 = 1$, then the process is called nonstationary process. In case of $a_1 = 1$ equity, the process is unit rooted and called random walk process. In order to present the differences between the unit rooted (nonstationary) series and the stationary series, two AR (1) process with sample sizes of 300 and with $(0,\sigma^2)$ distributed error terms, were produced. In the AR(1) process in Figure 2.1, while the a_1 coefficient is 0.8 and the process is stationary, in the AR(1) process in Figure 2.2, the a_1 coefficient is 1 and unit rooted, i.e. nonstationary.

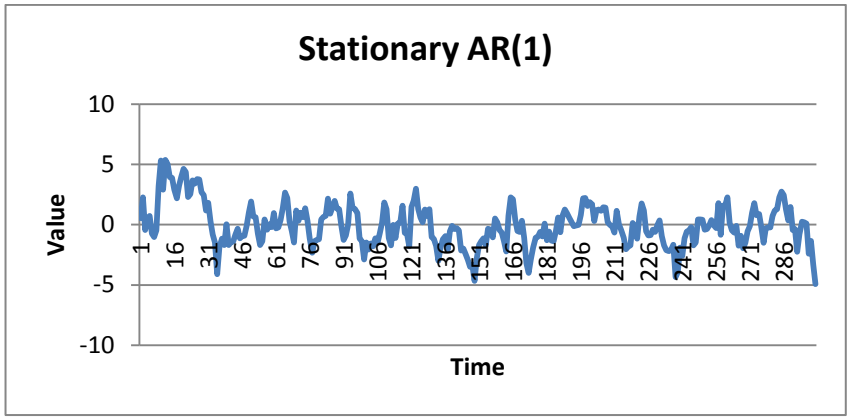

Figure 2.1 $x_t = 0.8x_{t-1} + \varepsilon_t$ Stationary AR (1) model

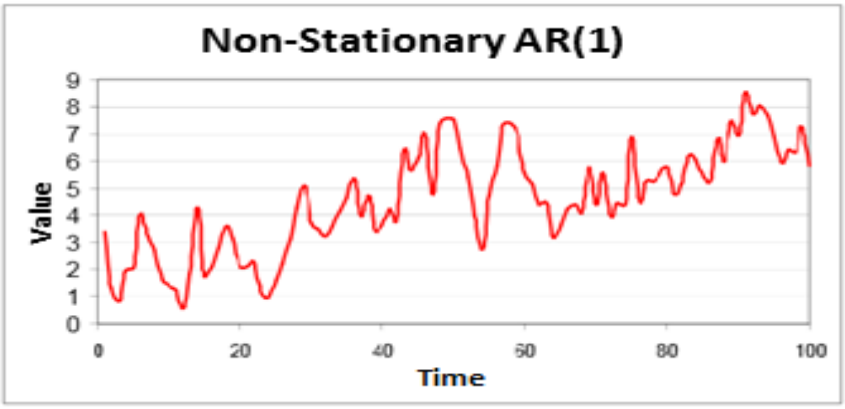

Figure 2.2 $x_t = x_{t-1} + \varepsilon_t$ Nonstationary AR (1) model (Random walk process)

Stationarity is a significant issue in a univariate time series as it is in multivariate time series. Let

$$\mathbf{X}_{it} \text{ i=1,2,…,k t=1,2,…,T}$$

be a k dimensional t time indexed time series variable. The stationarity conditions for the k dimensional $\mathbf{X}_{it}$ variable are as below;

a) If the $\mathbf{X}_{it}$ vector has time independent mean vector,

$$E(\boldsymbol{X_{it}}) = E(\boldsymbol{X_{is}}) = \mu_i \quad ,\forall \ t \neq s$$

b) If the covariance between the k dimensional $\mathbf{X}_{it}$ variable, which is realized in t time, and the $\mathbf{X}_{is}$ variable, which is realized in s time, depends only and only the time interval between these two variables, expressed as below:

$$Cov(\boldsymbol{X_{it}}, \boldsymbol{X_{is}}) = Cov\left(\boldsymbol{X_{i(t+j)}}, \boldsymbol{X_{i(s+j)}}\right) = \boldsymbol{\Gamma_s} \ , \forall \ j$$

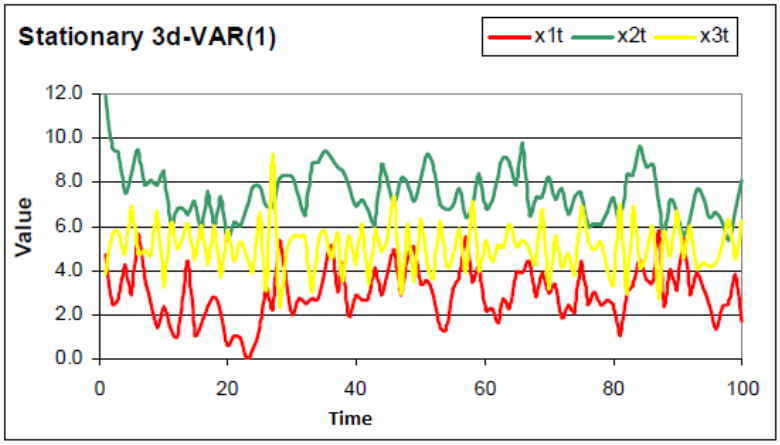

Figure 2.3 Three dimensional stationary VAR(1) process

The time series graph of the VAR (1) process, which has three variables, is presented in Figure 2.3. The equation of the process is given below:

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 5 \end{bmatrix} + \begin{bmatrix} 0.4 & 0.4 & 0.5 \\ 0 & 0.6 & 0 \\ 0 & 0 & -0.7 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix}$$

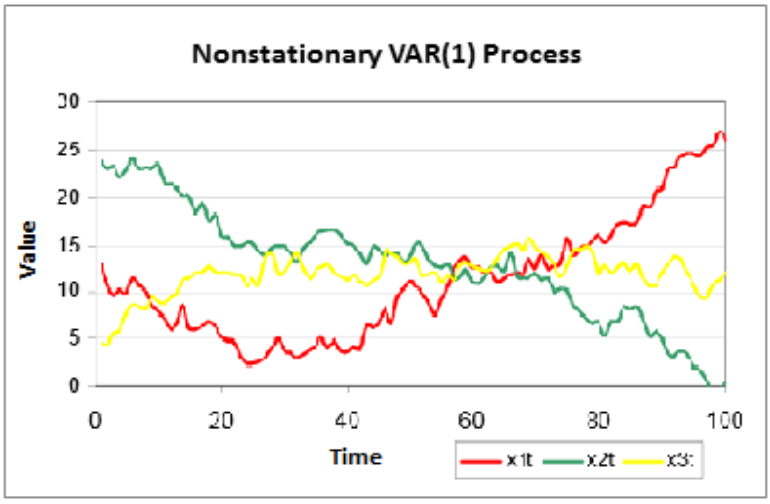

Figure 2.4 Three dimensional unrelated nonstationary VAR (1) process

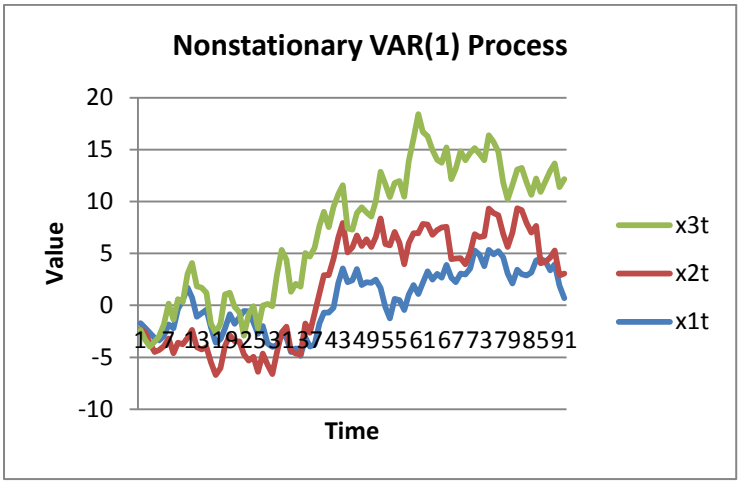

Figure 2.5 Three dimensional related nonstationary VAR (1) process

Figure 2.4 and Figure 2.5 presents two examples for nonstationary VAR (1) process with three variables. The models of these processes are given in Equations 2.2 and 2.3 respectively. Each variable in the model presented in Figure 2.4 were produced from the random walk process. In Figure 2.5, on the other hand, the processes X_{1t}, $X_{2t}$and X_{3t} were produced from the nonstationary random walk processes. While the variables in the model presented in Figure 2.4 are independent of each other, the first variable in the model in Figure 2.5 depends on the delay of other variables.

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} 20 \\ 30 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \tag{2.2}$$

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.4 & 0.4 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \tag{2.3}$$

The process in Figure 2.4 can be transformed into a stationary one by taking the difference of each series. On the contrary, the process in Figure 2.5 could not be transformed into a stationary one by taking the differences. This study aims at investigating the cointegrated relationships of the processes that cannot be smooth by difference operations.

2.2 Stationary VAR (p) Process

Before introducing the stationary VAR (p) process, brief information about the notation of the k dimensional $\mathbf{Y_t}$ vector and the matrix notation of the covariance matrix.

Let $\boldsymbol{Y_t}$vector present k variables observed in t time.

$$\mathbf{Y_t} = \begin{bmatrix} \mathbf{Y_{1t}} \\ \mathbf{Y_{2t}} \\ \vdots \\ \mathbf{Y_{kt}} \end{bmatrix} \quad -\infty < t < \infty$$

If k time series are observed in a certain time interval, $\boldsymbol{Y_t}$ vector for t=1,2,…,T could be expanded to a kxT dimension.

$$\mathbf{Y_t} = [\boldsymbol{Y_{1t}} \quad \boldsymbol{Y_{2t}} \quad \dots \quad \boldsymbol{Y_{kt}}]' = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1T} \\ y_{21} & y_{22} & \dots & y_{2T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{k1} & y_{k2} & \dots & y_{kT} \end{bmatrix}$$

Each row of **Y** matrix shows the univariate time series and the each column shows the values for the k^{th} variable in time t.

The covariance coefficient for the h^{th} step between the i^{th} and the j^{th} component of the $\mathbf{Y}_t$ vector is defined as below:

$$\Gamma_{ij}(h) = \mathrm{cov}\left(\mathbf{Y_{it}}, Y_{j,t-h}\right) = E(Y_{it} - \mu_i)(Y_{j,t-h} - \mu_j)$$

The kxk dimensional variance-covariance matrix of k variables for the h^{th} step is expressed as below:

$$\boldsymbol{\Gamma}(h) = E[\mathbf{Y_t} - \boldsymbol{\mu}][\mathbf{Y_{t-h}} - \boldsymbol{\mu}]' = \begin{bmatrix} \Gamma_{11}(h) & \Gamma_{12}(h) & \dots & \Gamma_{1k}(h) \\ \Gamma_{21}(h) & \Gamma_{22}(h) & \dots & \Gamma_{21}(h) \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{k1}(h) & \Gamma_{k2}(h) & \dots & \Gamma_{kk}(h) \end{bmatrix}$$

Let a_t shows the white noise process; the model of a p^{th} order autoregressive process AR(p) in univariate time series, is written as below:

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + a_t$$

In the multivariate time series model, on the other hand, the dependency structure of the i^{th} variable with other k variables. The model is as below:

$$Y_{it} = c_i + \emptyset_{i1,1}Y_{1t-1} + \emptyset_{i2,1}Y_{2t-1} + \cdots + \emptyset_{ik,1}Y_{kt-1} +$$

$$+\emptyset_{i1,2}Y_{1t-2} + \emptyset_{i2,2}Y_{2t-2} + \cdots + \emptyset_{ik,2}Y_{kt-2} +$$

$$+\emptyset_{i1,3}Y_{1t-3} + \emptyset_{i2,3}Y_{2t-3} + \cdots + \emptyset_{ik,3}Y_{kt-3} +$$

$$+ \cdots +$$

$$+\emptyset_{i1,p}Y_{1t-p} + \emptyset_{i2,p}Y_{2t-p} + \cdots + \emptyset_{ik,p}Y_{kt-p} + a_{it} \quad i = 1,2,\dots,k$$

The $\emptyset_{ik,p}$ parameter shows the i^{th} time series, k indicates the related variable, and p shows the delay degree of the model. The p^{th} order vector autoregressive process VAR (p) can be written with matrix notation as below:

$$\mathbf{Y_t} = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} + \begin{bmatrix} \emptyset_{11,1} & \emptyset_{12,1} & \dots & \emptyset_{1k,1} \\ \emptyset_{21,1} & \emptyset_{22,1} & \dots & \emptyset_{2k,1} \\ \vdots & \vdots & \vdots & \vdots \\ \emptyset_{k1,1} & \emptyset_{k2,1} & \dots & \emptyset_{kk,1} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{kt-1} \end{bmatrix} + \cdots +$$

$$+ \begin{bmatrix} \emptyset_{11,p} & \emptyset_{12,p} & \dots & \emptyset_{1k,p} \\ \emptyset_{21,p} & \emptyset_{22,p} & \dots & \emptyset_{2k,p} \\ \vdots & \vdots & \vdots & \vdots \\ \emptyset_{k1,p} & \emptyset_{k2,p} & \dots & \emptyset_{kk,p} \end{bmatrix} \begin{bmatrix} y_{1t-p} \\ y_{2t-p} \\ \vdots \\ y_{kt-p} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \\ \vdots \\ a_{kt} \end{bmatrix}$$

or

$$\boldsymbol{Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + a_t}$$

In the equation,

$\boldsymbol{Y_t}$: is the kx1 dimensional random time vector,

$\boldsymbol{\Phi_i}$: is the kxk dimensional autoregressive coefficients matrix i=1,2,…,p,

$\boldsymbol{c}$: is the kx1 dimensional constant term vector,

$\boldsymbol{a_t}$: is the kx1 dimensional white noise process vector.

The expected value and the variance-covariance matrices of the white noise process are as below:

$$E(\boldsymbol{a_t}) = \mathbf{0}$$

$$E(\boldsymbol{a_t a'_t}) = \boldsymbol{\Sigma_a} = \begin{bmatrix} E(a_{1t}^2) & E(a_{1t}a_{2t}) & \dots & E(a_{1t}a_{kt}) \\ E(a_{1t}a_{2t}) & E(a_{2t}^2) & \dots & E(a_{2t}a_{kt}) \\ \vdots & \vdots & \vdots & \vdots \\ E(a_{1t}a_{kt}) & E(a_{2t}a_{kt}) & \dots & E(a_{kt}^2) \end{bmatrix},$$

Where $\boldsymbol{\Sigma_a}$ is the positively defined variance-covariance matrix. Because of this property, $\boldsymbol{\Sigma_a}^{-1}$ is existing.

2.3 Stationary VAR (1) Process

After a brief information about the VAR (p) processes, the investigations will continue on the basis of VAR (1) process, in order to interpret the statistical properties of the process more easily.

2.3.1 Model Definition

1st order autoregressive process VAR (1) is presented in the equation below:

$$\boldsymbol{Y_t = c + \Phi_1 Y_{t-1} + a_t}$$

This model can be expressed using a backshift operator.

$$\boldsymbol{(I_k - \Phi_1 L) Y_t = c + a_t}$$

In the equation above the L is called as the backshift operator, and it shifts the variable it precedes backwards by the power of the operator ($\boldsymbol{L^j Y_t = Y_{t-j}}$).

2.3.2 General Linear Process

If we rewrite the VAR (1) model using the backshift method we obtain the equation below:

$$\boldsymbol{y_t = c + \Phi_1 y_{t-1} + a_t}$$

$$= \boldsymbol{c} + \emptyset_{\mathbf{1}}(\boldsymbol{c} + \ \boldsymbol{\Phi_1 y_{t-2}} + \boldsymbol{a_{t-1}}) + \boldsymbol{a_t}$$

$$= \mathbf{c} + \emptyset_{\mathbf{1}}\boldsymbol{c} + \boldsymbol{\Phi_1^2 y_{t-2}} + \boldsymbol{\Phi_1 a_{t-1}} + \boldsymbol{a_t}$$

$$= \mathbf{c} + \emptyset_{\mathbf{1}}\boldsymbol{c} + \emptyset_{\mathbf{1}}^{\mathbf{2}}(\boldsymbol{c} + \boldsymbol{\Phi_1 y_{t-3}} + \boldsymbol{a_{t-2}}) + \boldsymbol{\Phi_1 a_{t-1}} + \boldsymbol{a_t}$$

$$= \mathbf{c} + \emptyset_{\mathbf{1}}\boldsymbol{c} + \boldsymbol{\Phi_1^2 c} + \boldsymbol{\Phi_1^3 y_{t-3}} + \boldsymbol{\Phi_1^2 a_{t-2}} + \boldsymbol{\Phi_1 a_{t-1}} + \boldsymbol{a_t}$$

$$= (\boldsymbol{I} + \boldsymbol{\Phi_1} + \boldsymbol{\Phi_1^2})\boldsymbol{c} + \boldsymbol{\Phi_1^3 y_{t-3}} + \boldsymbol{\Phi_1^2 a_{t-2}} + \boldsymbol{\Phi_1 a_{t-1}} + \boldsymbol{a_t}$$

$$= \cdots$$

After performing n replacements VAR (1) model is expanded to the form below:

$$\boldsymbol{Y_t} = \left(\boldsymbol{I} + \boldsymbol{\Phi_1} + \boldsymbol{\Phi_1^2} + \cdots + \boldsymbol{\Phi_1^n}\right)\boldsymbol{c} + \boldsymbol{\Phi_1^{n+1} Y_{t-n-1}} + \boldsymbol{\Phi_1^n a_{t-n}} + \cdots + \boldsymbol{\Phi_1 a_{t-1}} + \boldsymbol{a_t}$$

As the VAR (1) model will become close to zero, when the $\boldsymbol{\Phi_1^{n+1}}$ coefficient of $\boldsymbol{y_{t-n-1}}$ variable is n → ∞, and thus become negligible, VAR (1) becomes a model which comprises only of random shocks. Therefore, VAR (1) model, which has the property of invertibility in its nature, becomes a stationary model. If it is accepted that the kxk dimensional matrix with the $\boldsymbol{\Phi_1}$ coefficient has s linear independent eigenvectors, s being less than or equal to k, the $\boldsymbol{\Phi_1}$ matrix is tried to be decomposed using the Jordan Decomposition method. P_{kxk} being any matrix that is not singular, the equation below can be written:

$$\boldsymbol{\Phi_1} = \boldsymbol{P\Lambda P^{-1}}$$

The Λ matrix with the diagonal elements being Λ_i and other elements being zero is presented as below:

$$\boldsymbol{\Lambda}: = \begin{bmatrix} \boldsymbol{\Lambda_1} & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \boldsymbol{\Lambda_n} \end{bmatrix}$$

The $\boldsymbol{\Lambda_i}$ matrix, with the diagonal elements comprising of the eigenvalues (λ_i i=1,2,…,k) obtained from the $\boldsymbol{\Phi_1}$ matrix, and with other elements, crossover the diagonal elements, being zero, is written as below:

$$\boldsymbol{\Lambda_i} = \begin{bmatrix} \lambda_i & 1 & 0 & \dots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ 0 & & \ddots & \ddots & \vdots \\ \vdots & 0 & & \ddots & 1 \\ 0 & 0 & \dots & \dots & \lambda_i \end{bmatrix}$$

Then, the m^{th} power of the $\boldsymbol{\Phi_1}$ matrix, decomposed using the Jordan Decomposition method, can easily be obtained by using the equation below, with the help of the decomposed matrices:

$$\boldsymbol{\Phi_1^m = (PJP^{-1})^m = PJ^mP^{-1}}$$

If the absolute values of the eigenvalues of the $\boldsymbol{\Phi_1}$ matrix are less than 1, for $m \to \infty$, $\boldsymbol{\Phi_1^m} = PJ^mP^{-1} \to 0$ will be close.Since it would converge to $\boldsymbol{\Phi_1^{n+1} y_{t-n-1} \to 0}$ value $\Phi_1^{n+1} y_{t-n-1}$ can be neglected in the model. Therefore VAR (1) model is stationary (Hamilton 1994).

2.4 Near-Stationary Process

This section will cover the process which are close to unit root, except a unit root case. This process which was introduced into the literature by Philips (1987b) is widely used in the application; since the financial or economic data are either unit rooted or have parameters close to unit root.

If we consider the AR (1), it is written as below:

$$y_t = \rho y_{t-1} + u_t \tag{2.4}$$

Here u_t is independent and has $N(0, \sigma^2)$ distribution. When $|\rho| < 1$ and $y_0 \sim N(0, \sigma^2(1-\rho^2)^{-1})$,$\{y_t\}$ is a stationary process. On the other hand, if $\rho = 1$ and $y_0 = 0$, $\{y_t\}$ is first order integrated (I(1)) and nonstationary; the variance of the series varies depending on t, i.e. the variance is equal to the value $t\sigma^2$.

In case $|\rho| < 1$ and very close to 1; i.e. for ε<0, when the ρ value is defined as $\rho = 1 + \varepsilon$ for the small ε values, the equation will be as below:

$$E(y_t) = 0$$

$$Var(y_t) = \sigma^2[1 + (1+\varepsilon)^2 + \cdots + (1+\varepsilon)^{2(t-1)}]$$

$$= \sigma^2 t(1 + \varepsilon(t-1)) \approx \sigma^2 t$$

Therefore, the variance would behave as a trend equation. In other words, the process would behave as a first order integrated process although it is an asymptotically stationary.

A more appropriate parameterisation could be used for a near-stationary process:

$$\rho = \exp\left(\frac{\varepsilon}{T}\right) \qquad \varepsilon < 0$$

This parameterisation will create a local alternative series for $\rho = 1$ series. When $\varepsilon = 0$, it will result in $\rho = 1$, and ρ will obtain a value less than 1 but very close to 1 for the smaller ε values, if ε<0. Put it another way, for $\to \infty$, it will converge to $\rho \to 1$ value. Thus, the process will be called as near-integrated process; because the process, for the smaller ε values, for ε<0, will behave as a first order integrated process.

2.5 Non-stationary Time Series

Nonstationary time series would yield unit rooted parameters in the prediction and regression equations. It will be give brief information about the nonstationary, in this section univariate and multivariate time series models which include unit roots.

2.5.1 Unit Root Status in Univariate Time Series

The situation that the univariate time series has a unit root means that the series under investigation is nonstationary. Series like this contains a stochastic trend. If we are to consider the AR (1) model

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

in case of $\rho = 1$, the parameter is unit rooted and the process is nonstationary. As it is mentioned in the previous sections, processes like these are called random walk process. When the errors are assumed as zero mean, with unit variance and normal

distribution and when initial value is $y_0 = 0$, the random walk process will be obtained as below:

$$y_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$

$$y_t \sim N(0, t)$$

The variance in y at time t and s is below,

$$y_s - y_t = \varepsilon_{t+1} + \varepsilon_{t+2} + \cdots + \varepsilon_s$$

$$y_s - y_t \sim N(0, (s - t))$$

The variance of the y_t process at the time interval between s and t is distributed normally. This situation will be investigated in detail in the following chapters. The processes which are nonstationary, such as random walk process, will be divided into parts based on equal intervals, and the distribution of the divided parts will be obtained using the Wiener process.

2.5.2 Unit Root Situation in Multivariate Time Series

As it is in univariate processes, unit root situation is encountered in multivariate processes. If the first order autoregressive process with variable vector is considered, the model of the process will be as below:

$$\boldsymbol{X_t = A_1 X_{t-1} + \varepsilon_t}$$

In the model the equity of the $\boldsymbol{A_1}$ coefficient matrix to the unit matrix indicates that the model includes a unit root. In this situation the equation of the model will become as below:

$$\boldsymbol{X_t = X_{t-1} + \varepsilon_t}$$

If at least one of the elements of the nonstationary multivariate time vector $(X_{1t}, \ldots, X_{kt})$ includes a unit root, the process will not have the property of stationarity.

CHAPTER THREE
THE ASYMPTOTIC PROPERTIES OF UNIT ROOT PROCESSES

This chapter aims give information about the asymptotic distribution of unit rooted processes and the properties of these distributions. For this purpose, first the Brownian motion functions and some theorems should be considered.

3.1 Brownian Motion (Wiener process)

In statistics, the Wiener process is a continuous-time stochastic process named in Norbert Wiener. It is often called standard Brownian motion, after Robert Brown.

3.1.1 Random Walk and Brownian Motion

It was previously mentioned in Chapter Two that the random walk in Equation 3.1 is not a stationary process.

$$y_t = y_{t-1} + a_t \tag{3.1}$$

When the process is started to be written with $y_0 = 0$ assuming that the errors in the random walk process are with zero mean and unit variance normal distribution "$a_t \sim i.i.d\ N(0,1)$", y_t will be obtained as below.

$$y_t = a_1 + a_2 + \cdots + a_t$$

$$y_t \sim N(0, t)$$

The change in y in times t and s will be as follows,

$$y_s - y_t = a_{t+1} + a_{t+2} + \cdots + a_s$$

$$y_s - y_t \sim N(0, (s-t))$$

and this change will be independent of any changes in r and q times (assuming t<s<r<q). When the change of y in time t-1 and t is examined by dividing the t-1 and t time period to n sub-periods, the process will become as follows:

$$y_t - y_{t-1} = a_{1t} + a_{2t} + \cdots + a_{nt}$$

Here it should be noted that $a_{it} \sim i.i.d.\ N(0,1/n)$. This process has the same properties with the random walk process. The limit of this process for $n \to \infty$ will return the continuous times process known as the standard Brownian motion. The value of this process in t time is shown with W(t).

Brownian motion will be discussed in detail under the functional central limit theorem.

3.1.2 Standard Brownian Motion (Standard Wiener Process)

The use of continuous timed stochastic processes to obtain the asymptotic processes of the unit rooted processes is quite frequent. Wiener process is used for describing of the asymptotic properties of the estimated parameters. Standard Brownian motion or standard wiener process are examples of the continuous timed processes in the range of (W(.)), [0,1].

For each t$\in$ [0,1], W(t) has the three properties below.

(a)W(0)=0;

(b) Considering $W(r) \sim N(0,r)$, for
$0 \le t_1 \le t_2 \le \cdots \le t_k$ $[W(t_2) - W(t_1)], \dots, [W(t_k) - W(t_{k-1})]$ random variables are independent form each other and normally distributed.

(c) W(t) is a continuous function of t in almost everywhere.

In order to develop the unit root asymptotics, generally the $X_T(r)$ quantity is considered. Here

$$X_T(r) \coloneqq T^{-1} \sum_{t=1}^{[\![Tr]\!]} a_t \qquad r \epsilon [0,1]$$

a_t represents a stationary stochastic process, r $\in$ [0,1] represents the proportion, $[\![Tr]\!]$ represents the integer value of Tr expression. Use of $X_T(r)$ quantity in unit rooted processes will be explained in detail when discussing the functional central limit theorem.

3.2 Functional Central Limit Theorem

Using the functional central limit theorem, it is tried to prove, by dividing the unit rooted processes into different sub-period intervals, that the functions belonging to these intervals have independent and identical distributions with different parameters.

As per the central limit theorem, and considering u_t as being independent random variables having the same distribution with a zero mean and σ^2 variance, the asymptotic distribution of $\bar{u}_t$ sample mean will have a normal distribution as shown in Equation 3.3.

$$\bar{u}_t = T^{-1} \sum_{t=1}^{T} a_t \tag{3.2}$$

$$\sqrt{\boldsymbol{T}}\bar{u}_t \xrightarrow{d} N(0, \sigma^2) \tag{3.3}$$

Let us try to obtain the $X_T(r)$ step function by dividing the sample into certain subsections using the function pertaining to the sample mean in order to show the functions of the functional central limit theorem.

$$X_T(r) \coloneqq T^{-1} \sum_{t=1}^{[\![Tr]\!]} a_t \qquad r\epsilon[0,1]$$

The step function of $X_T(r)$is as below:

$$X_T(r) = \begin{cases} 0 & 0 \le r < \dfrac{1}{T} \\ \dfrac{a_1}{T} & \dfrac{1}{T} \le r < \dfrac{2}{T} \\ \vdots & \\ \dfrac{(a_1 + a_2 + \cdots + a_T)}{T} & r = 1 \end{cases}$$

Later, for all values of r, as per the central limit theorem, it is obvious that it would be;

$$\sqrt{\boldsymbol{T}} X_T(r) = \frac{1}{\sqrt{\boldsymbol{T}}} \sum_{t=1}^{[\![Tr]\!]} a_t = \left(\frac{\sqrt{[\![Tr]\!]}}{\sqrt{T}}\right)\left(\frac{1}{\sqrt{[\![Tr]\!]}}\right) \sum_{t=1}^{[\![Tr]\!]} a_t$$

$$\left(\frac{1}{\sqrt{[\![Tr]\!]}}\right)\sum_{t=1}^{[\![Tr]\!]} a_t \xrightarrow{L} N(0,\sigma^2)$$

Since $\left(\frac{\sqrt{[\![Tr]\!]}}{\sqrt{T}}\right) \to \sqrt{r}$,

$$\sqrt{\boldsymbol{T}}X_T(r) \xrightarrow{d} N(0, r\sigma^2)$$

$$\sqrt{\boldsymbol{T}}(X_T(r)/\sigma) \xrightarrow{d} N(0, r)$$

they will have the distributions above.

Similarly, let $r_2 > r_1$, the asymptotic distribution, mean and variance of the difference of the two step functions for a $[\![Tr_1]\!]$ and a $[\![Tr_2]\!]$ sample size would fit the normal distribution as in Equation 3.4.

$$\sqrt{\boldsymbol{T}}(X_T(r_2) - X_T(r_1))/\sigma \xrightarrow{L} N(0, r_2 - r_1) \tag{3.4}$$

According to all these results, it will have the $\sqrt{\boldsymbol{T}}(X_T(.)/\sigma)$ properties

which takes the $X_T(r)$=0 for r=0 ,

which takes continuous values for each t and

asymptotic normally distributed for each $r_2 > r_1$value.

$$\sqrt{\boldsymbol{T}}(X_T(.)/\sigma) \coloneqq W(.)$$

The Brownian motion provides the possibility of applying the central limit theorem in a more general way. Using the functional central limit theorem, the traditional central limit theorem for r=1 can easily be reached. In other words;

$$X_T(1) = T^{-1}\sum_{t=1}^{T} a_t \qquad r = 1$$

$$\sqrt{\boldsymbol{T}}(X_T(.)/\sigma) = [1/\sigma\sqrt{T}\sum_{t=1}^{T} a_t] \xrightarrow{d} W(1) \sim N(0,1)$$

is obtained

3.3 Continuous Mapping Theorem

Let $\{X_T\}_{t=1}^{\infty}$random variables array be a $g: R \to R$ defined continuous function and let X be any random variable.As per the continuous mapping theorem, while $T \to \infty$, if $X_T \xrightarrow{d} X$ approximates, $g(X_T) \xrightarrow{d} g(X)$ will approximate.Accordingly, while $T \to \infty$,

$$\sqrt{T}(X_T(.)) \coloneqq \sigma W(.) \quad , W(r) \sim N(0,r) \quad ve\ \sqrt{T}X_T(r) \xrightarrow{d} \sigma W(r)$$

can be written.Also if $S_T(.)$Is a stochastic function of X random variable (i.e.y=$\int_0^1 S(r)dr$), it is obvious that $S_T(r) = [\sqrt{T}X_T(r)]^2 \quad , \quad S_T(r) \xrightarrow{d} \sigma^2[W(r)]^2$.

3.4 Asymptotic Properties for Unit Root Univariate Time Series

Philips (1986, 1987) is the first researcher to investigate the asymptotic distributions of the unit rooted processes using the functional central limit theorem. In this section, the obtaining process of the asymptotic distributions of the unit rooted AR (1) process will be summarized. The equation of the unit rooted AR (1) process is presented below:

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

For the random walk process if $a_t$is accepted as independent and having the same distribution with zero mean and σ^2 variance and if the process starts with $y_0 = 0$, the model will be as

$$y_t = a_1 + \cdots + a_t \tag{3.5}$$

The model shown in Equation (3.5), $X_T(r)$ step function is as below:

$$X_T(r) = \begin{cases} 0 & 0 \le r < \frac{1}{T} \\ \frac{y_1}{T} & \frac{1}{T} \le r < \frac{2}{T} \\ \vdots & \\ \frac{y_T}{T} & r = 1 \end{cases}$$

The graph of the step function obtained above is given in Figure 3.1

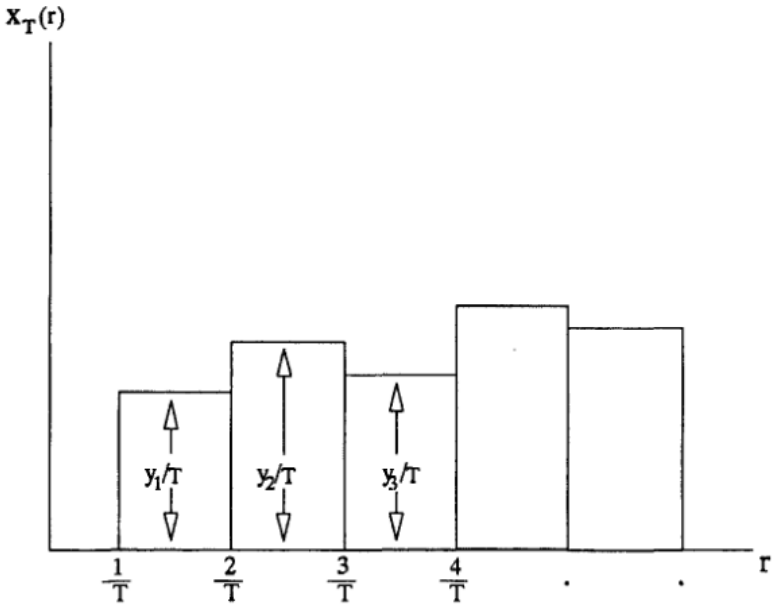

Figure 3.1 Graph of $X_T(r)$step function

When figure 3.1 is examined; the area below the function will be equal to the product of T rectangles. In other words, the area of the first rectangle is obtained by the product of the length of base $1/T$ and the height $\frac{y_1}{T}$. When all the areas of the rectangles are summed, the area of the $X_T(r)$ step function will between [0,1] interval. The mathematical representation is given in (3.6).

$$\int_0^1 X_T(r)\,dr = \frac{y_1}{T^2} + \cdots + \frac{y_{T-1}}{T^2} = \sum_{t=1}^{T} \frac{y_{t-1}}{T^2} \tag{3.6}$$

When Equation (3.6) is multiplied by $\sqrt{T}$, the product is:

$$\int_0^1 \sqrt{T} X_T(r)\,dr = T^{-3/2} \sum_{t=1}^{T} y_{t-1} \tag{3.7}$$

Equation (3.7), will become Wiener process for $T \to \infty$ as per continuous mapping theorem.

$$\int_0^1 \sqrt{T} X_T(r)\,dr \xrightarrow{d} \sigma \int_0^1 W(r)\,dr \tag{3.8}$$

The information in Equation (3.8), can be used to obtain the asymptotic distribution of the $T^{-3/2}\sum_{t=1}^{T} y_{t-1}$ function obtained in (3.7).

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{L} \sigma \int_0^1 W(r)\, dr$$

When the equation $T^{-3/2}\sum_{t=1}^{T} y_{t-1}$ is unfolded until T equation (3.9) is obtained.

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1} = T^{-\frac{3}{2}}[a_1 + (a_1 + a_2) + (a_1 + a_2 + a_3) + \cdots + (a_1 + a_2 + a_3 + \cdots + a_{T-1})]$$

$$= T^{-3/2}\big[(T-1)a_1 + (T-2)a_2 + (T-3)a_3 + \cdots + \big(T-(T-1)\big)a_{T-1}\big] = T^{-\frac{3}{2}}\sum_{t=1}^{T}(T-t)a_t$$

$$= T^{-1/2}\sum_{t=1}^{T} a_t - T^{-\frac{3}{2}}\sum_{t=1}^{T} t a_t \qquad (3.9)$$

Hamilton (1994) showed that the function in (3.9) approximated the normal distribution, with the parameters given below.

$$\begin{bmatrix} (\frac{1}{\sqrt{T}})\sum a_t \\ (\frac{1}{\sqrt{T}})\sum (t/T)a_t \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \sigma^2\begin{bmatrix}1 & 1/2\\ 1/2 & 1/3\end{bmatrix}\right)$$

Thus $T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1}$ asymptotically will approximate to the normal distribution with 0 mean and $\frac{\sigma^2}{3}$ variance. Using the equation in (3.11), $\sigma\int_0^1 W(r)\, dr$ expression will naturally have the same distribution; in other words it will also be N(0,$\frac{\sigma^2}{3}$). With reference to this information, the asymptotic distribution of the function $T^{-\frac{3}{2}}\sum_{t=1}^{T} t a_t$ will be as given in Equation (3.10).

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} ta_t = T^{-\frac{1}{2}}\sum_{t=1}^{T} a_t - T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1}$$

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} ta_t \xrightarrow{L} \sigma W(1) - \sigma\int_0^1 W(r)\,dr \qquad (3.10)$$

The asymptotic distribution of the sum of stochastic squares of the random walk process can be obtained similarly. If $S_T(r)$ statistics is defined as below;

$$S_T(r) = [\sqrt{T}X_T(r)]^2$$

The step function $S_T(r)$ statistics can be obtained as below:

$$S_T(r) = \begin{cases} 0 & 0 \le r < \frac{1}{T} \\ \frac{{y_1}^2}{T} & \frac{1}{T} \le r < \frac{2}{T} \\ \vdots & \\ \frac{{y_T}^2}{T} & r = 1 \end{cases}$$

Later; The area of $S_T(r)$ in the [0,1] interval will be equal to equation $\sum_{t=1}^{T} \frac{{y_{t-1}}^2}{T^2}$.

$$\int_0^1 S_T(r)\,dr = \frac{{y_1}^2}{T^2} + \cdots + \frac{{y_{T-1}}^2}{T^2} = \sum_{t=1}^{T} \frac{{y_{t-1}}^2}{T^2}$$

As per the continuous mapping theorem;

$$T^{-2}\sum_{t=1}^{T} {y_{t-1}}^2 \xrightarrow{d} \sigma^2\int_0^1 [W(r)]^2 dr$$

can be written.When the asymptotic distribution of the products of the statistics discussed up to now with t/T, assuming r=t/T, equation (3.11) is obtained.

$$T^{-5/2}\sum_{t=1}^{T} ty_{t-1} = T^{-3/2}\sum_{t=1}^{T} (\frac{t}{T})y_{t-1} \xrightarrow{d} \sigma\int_0^1 rW(r)\,dr \qquad (3.11)$$

Similarly;

$$T^{-3}\sum_{t=1}^{T} ty_{t-1}^2 = T^{-2}\sum_{t=1}^{T}(\frac{t}{T})y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 r[W(r)]^2 dr$$

Until now, the asymptotic distributions of $y_t$and its different functions have been investigated. Using the results obtained, one can get information about the asymptotic distribution of unit rooted AR (1) model. The operation begins by squaring both sides of the AR (1) model. Let

$$y_t^2 = (y_{t-1} + a_t)^2 = y_{t-1}^2 + 2y_{t-1}a_t + a_t^2$$

$$y_{t-1}a_t = \left(\frac{1}{2}\right)\{y_t^2 - y_{t-1}^2 - a_t^2\}$$

and for $y_0 = 0$;

$$T^{-1}\sum_{t=1}^{T} y_{t-1}a_t = \left(\frac{1}{2}\right)(1/T)\{y_T^2\} - \left(\frac{1}{2}\right)(\frac{1}{T})\sum_{t=1}^{T} a_t^2$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1}a_t = \left(\frac{1}{2}\right)S_T(1) - \left(\frac{1}{2}\right)\left(\frac{1}{T}\right)\sum_{t=1}^{T} a_t^2$$

are obtained. Here, as per the law of large numbers and the continuous mapping theorem, the distributions will be obtained:

$$(\frac{1}{T})\sum_{t=1}^{T} a_t^2 \xrightarrow{p} \sigma^2$$

$$S_T(1) \xrightarrow{d} \sigma^2[W(1)]^2$$

Here W(1), is a wiener process with N(0,1) distribution. With regard to this information $[W(1)]^2$ will distribute as $\chi^2(1)$ and will be equal to the expression below:

$$T^{-1}\sum_{t=1}^{T} y_{t-1}a_t \xrightarrow{d} \left(\frac{1}{2}\right)\sigma^2[W(1)]^2 - \left(\frac{1}{2}\right)\sigma^2 = \left(\frac{1}{2}\right)\sigma^2[[W(1)]^2 - 1]$$

$$= \left(\frac{1}{2}\right)\sigma^2[\chi^2(1) - 1]$$

The asymptotic properties for Univariate Unit Rooted AR(1) process are summarized in Table 3.1.

Table 3.1 The asymptotic distributions of wiener processes

Functions	The asymptotic distributions of wiener processes used
$\left(\frac{1}{\sqrt{T}}\right)\sum a_t \xrightarrow{d} N(0,\sigma^2)$	
$T^{-1}\sum_{t=1}^{T} y_{t-1}a_t \xrightarrow{d} \left(\frac{1}{2}\right)\sigma^2[[W(1)]^2 - 1]$	$W(1)^2 \sim \mathcal{X}^2_{(1)}$
$T^{-\frac{3}{2}}\sum_{t=1}^{T} ta_t \xrightarrow{d} \sigma W(1) - \sigma\int_0^1 W(r)\,dr$	W(1)~ N(0,1) $\int_0^1 W(r)\,dr$ ~N(0,1/3)
$T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma\int_0^1 W(r)\,dr$	$\int_0^1 W(r)\,dr$ ~N(0,1/3)
$T^{-2}\sum_{t=1}^{T} {y_{t-1}}^2 \xrightarrow{d} \sigma^2\int_0^1 [W(r)]^2 dr$	$[W(r)]^2 \sim \mathcal{X}^2_{(r)}$
$T^{-5/2}\sum_{t=1}^{T} ty_{t-1} \xrightarrow{d} \sigma\int_0^1 rW(r)\,dr$	

3.5 Asymptotic Properties for Unit Root Multivariate Time Series

In the previous section the asymptotic properties of univariate AR(1) process and naturally the univariate standard Brownian motion W(r) were discussed. Since the asymptotic properties of the multivariate VAR(1) process will be discussed in this section, information about the multivariate Brownian motion **W(r)** will be provided. **W(r)** is a (nx1) sized vector which contains n "$W_1(r), \dots, W_n(r)$" processes independent from each other.

The n sized **W(r)** multivariate standard Brownian motion defined $r \in [0,1]$ has the three properties given below.

a) $\boldsymbol{W(0) = 0}$

b) For any $0 \le r_1 < r_2 < \cdots < r_k < 1$ time, $[\boldsymbol{W(r_2) - W(r_1)}], \dots, [\boldsymbol{W(r_k) - W(r_{k-1})}]$ changes have independent and identical distributions. They are normally distributed as $[\boldsymbol{W(s) - W(r)}] \sim N(\mathbf{0}, (s-r)\boldsymbol{I_n})$.

c) **W(r),** is a continuous stochastic function with a probability of 1 in the interval in which r is defined.

Let the $\{v_t\}_{t=1}^{\infty}$univariate discrete time process with 0 mean and unit variance, which are independent and have the same distribution, be defined.

$$X_T^*(r) \equiv T^{-1}[v_1 + v_2 + \cdots + v_{[Tr]*}]$$

$[Tr]*$ expresses the integral function which shows the largest integer value which is equal to or smaller than Tr. As per functional central limit theorem, for $T \to \infty$ it was;

$$\sqrt{T} X_T^*(.) \xrightarrow{d} W(.)$$

This situation written for univariate processes can be generalized for multivariate processes.

For the multivariate $\{\boldsymbol{v_t}\}_{t=1}^{\infty}$ vector process with independent variables and the same distribution let $E(\boldsymbol{v_t}) = \mathbf{0}$ and $E(\boldsymbol{v_t v_t'}) = \boldsymbol{I_n}$, for the function below

$$\boldsymbol{X_T^*(r)} = T^{-1}[\mathbf{v_1} + \mathbf{v_2} + \cdots + \mathbf{v}_{[\mathbf{Tr}]*}] \qquad (3.12)$$

It is written as below

$$\sqrt{T}\boldsymbol{X}_{\boldsymbol{T}}^{*}(.) \xrightarrow{d} \boldsymbol{W}(.)$$

Later, if the $\{\boldsymbol{\varepsilon}_{\boldsymbol{t}}\}_{t=1}^{\infty}$ process, with n-size, zero mean vector and $\boldsymbol{\Omega}$ variance - covariance matrix, is defined the $\mathbf{P}$ matrix can be obtained by applying cholesky decomposition for the $\boldsymbol{\Omega}$ variance-covariance matrix.

$$\boldsymbol{\Omega} = \mathbf{P}\mathbf{P}'$$

If it is assumed that all $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$ are produced from $\boldsymbol{v}_{\boldsymbol{t}}$ the equation below can be written by using the P matrix obtained from the decomposition.

$$\boldsymbol{\varepsilon}_{\boldsymbol{t}} = \mathbf{P}\boldsymbol{v}_{\boldsymbol{t}}$$

When $\boldsymbol{\varepsilon}_{\boldsymbol{t}}$ is expressed with $\mathbf{P}\boldsymbol{v}_{\boldsymbol{t}}$ its initial properties do not change. Its mean and variance-covariance matrix are as below:

$$E(\boldsymbol{\varepsilon}_{\boldsymbol{t}}) = \mathbf{P}\mathrm{E}(\boldsymbol{v}_{\boldsymbol{t}}) = \boldsymbol{P} * \mathbf{0} = \mathbf{0}$$

$$E(\boldsymbol{\varepsilon}_{\mathbf{t}}\boldsymbol{\varepsilon}_{\mathbf{t}}') = \mathbf{P}E(\boldsymbol{v}_{\boldsymbol{t}}\boldsymbol{v}_{\boldsymbol{t}}')\mathbf{P}' = \mathbf{P}\mathbf{I}_{\mathbf{n}}\mathbf{P}' = \boldsymbol{\Omega}$$

The properties of the $\boldsymbol{X}_{\boldsymbol{T}}^{*}(\boldsymbol{r})$ statistics given in Equation (3.12) will be as below:

$$\boldsymbol{X}_{\boldsymbol{T}}^{*}(\boldsymbol{r}) = T^{-1}[\boldsymbol{\varepsilon}_{\mathbf{1}} + \boldsymbol{\varepsilon}_{\mathbf{2}} + \cdots + \boldsymbol{\varepsilon}_{[\mathbf{Tr}]*}]$$

$$= \mathbf{P}\mathrm{T}^{-1}[\mathbf{v}_{\mathbf{1}} + \mathbf{v}_{\mathbf{2}} + \cdots + \mathbf{v}_{[\mathbf{Tr}]*}]$$

$$= \mathbf{P}\boldsymbol{X}_{\boldsymbol{T}}^{*}(\boldsymbol{r})$$

As per the continuous mapping theorem, it will distribute as below:

$$\sqrt{\mathrm{T}}\mathbf{X}_{\mathbf{T}}^{*}(.) \xrightarrow{\mathrm{d}} \mathbf{P}\mathbf{W}(.)$$

For any r, $\mathbf{PW}(\mathbf{r})$ has N $(\mathbf{0}, \mathbf{r\Omega})$ distribution.

The $\mathbf{a}_{\mathbf{t}}$ vector in the multivariate VAR (1) process can be written as dependent $\boldsymbol{\varepsilon}_{\mathbf{t}}$ vectors using the functional central limit theorem.

$$\mathbf{a_t} = \sum_{s=0}^{\infty} \mathbf{\Psi_s \varepsilon_{t-s}}$$

Let $\mathbf{\Psi}_{ij}$ be the ith line and jth column element of the matrix $\mathbf{\Psi_s}$. The lineer weights of $\mathbf{a_t}$ obtained as $\mathbf{\varepsilon_t}$ vectors should be smaller than infinite for each i and j, as per the stability and convertibility condition. This condition is stated below:

$$\sum_{s=0}^{\infty} s\left|\mathbf{\Psi}_{ij}\right| < \infty$$

Since the unit root VAR (1) models are under discussion, at this point it is necessary to give information about the Beveridge-Nelson decomposition. Beveridge-Nelson decomposition is used for dividing a scalar time series vector $\mathbf{Y_t}$ into two parts as total random walk component and stationary component.Thus, information about the asymptotic properties of the unit root component can be obtained.

Beveridge-nelson decomposition shows that an I (1) process can be written as the initial conditions of the sum of random walks and as the linear components of a stationary process. The way to obtain the decomposition is given in Equation 3.13 from its beginning to the end.

$$\sum_{s=1}^{t} \mathbf{a_s} = \mathbf{a_1} + \cdots + \mathbf{a_t} = \sum_{s=0}^{\infty} \mathbf{\Psi_s \varepsilon_{1-s}} + \sum_{s=0}^{\infty} \mathbf{\Psi_s \varepsilon_{2-s}} + \cdots + \sum_{s=0}^{\infty} \mathbf{\Psi_s \varepsilon_{t-s}}$$

$$= (\mathbf{\Psi_0 \varepsilon_1} + \mathbf{\Psi_1 \varepsilon_0} + \mathbf{\Psi_2 \varepsilon_{-1}} + \mathbf{\Psi_3 \varepsilon_{-2}} + \cdots)$$

$$+(\mathbf{\Psi_0 \varepsilon_2} + \mathbf{\Psi_1 \varepsilon_1} + \mathbf{\Psi_2 \varepsilon_0} + \mathbf{\Psi_3 \varepsilon_{-1}} + \cdots)$$

$$\vdots$$

$$+(\mathbf{\Psi_0 \varepsilon_t} + \mathbf{\Psi_1 \varepsilon_{t-1}} + \mathbf{\Psi_2 \varepsilon_{t-2}} + \mathbf{\Psi_3 \varepsilon_{t-3}} + \cdots)$$

$$= \mathbf{\Psi_0}(\mathbf{\varepsilon_1} + \cdots + \mathbf{\varepsilon_t}) + \mathbf{\Psi_1}(\mathbf{\varepsilon_1} + \cdots + \mathbf{\varepsilon_t}) + \cdots$$

$$-(\mathbf{\Psi_1} + \mathbf{\Psi_2} + \cdots)\mathbf{\varepsilon_t} - (\mathbf{\Psi_2} + \mathbf{\Psi_3} + \cdots)\mathbf{\varepsilon_{t-1}} - (\mathbf{\Psi_3} + \mathbf{\Psi_4} + \cdots)\mathbf{\varepsilon_{t-2}} - \cdots$$

$$= (\boldsymbol{\Psi_0} + \boldsymbol{\Psi_1} + \boldsymbol{\Psi_2} + \cdots)(\boldsymbol{\varepsilon_1} + \cdots + \boldsymbol{\varepsilon_t}) - (\boldsymbol{\Psi_1} + \boldsymbol{\Psi_2} + \cdots)\boldsymbol{\varepsilon_t}$$
$$- (\boldsymbol{\Psi_2} + \boldsymbol{\Psi_3} + \cdots)\boldsymbol{\varepsilon_{t-1}} - (\boldsymbol{\Psi_3} + \boldsymbol{\Psi_4} + \cdots)\boldsymbol{\varepsilon_{t-2}} - \cdots$$

$$\sum_{s=1}^{t} \mathbf{a_s} = \boldsymbol{\Psi}(\mathbf{1}) \sum_{s=1}^{t} \boldsymbol{\varepsilon_s} + \boldsymbol{\eta_t} - \boldsymbol{\eta_0} \tag{3.13}$$

In the equation above, $\boldsymbol{\Psi}(\mathbf{1}) = (\boldsymbol{\Psi_0} + \boldsymbol{\Psi_1} + \boldsymbol{\Psi_2} + \cdots)$ and $\boldsymbol{\eta_t} = \sum_{\mathbf{s=0}}^{\infty} \boldsymbol{\alpha_s}\boldsymbol{\varepsilon_{t-s}}$ are in the form of $\boldsymbol{\alpha_s} = -(\boldsymbol{\Psi_{s+1}} + \boldsymbol{\Psi_{s+2}} + \cdots)$. The term $\boldsymbol{\Psi}(\mathbf{1}) \sum_{s=1}^{t} \boldsymbol{\varepsilon_s}$ in equation 3.13 shows the random walk part, the $\boldsymbol{\eta_t}$ term shows the stationary part and $\mathbf{y_0} - \boldsymbol{\eta_0}$ shows the initial conditions.

The step function, used in investigating the asymptotic properties for the univariate unit root time series, can be used with the same purpose for multivariate unit root time series. For the equation

$$\mathbf{X_T}(\mathbf{r}) = (\frac{1}{T}) \sum_{s=1}^{[\mathbf{Tr}]*} \mathbf{a_s}$$

The equation below is written using Beveridge-Nelson decomposition,

$$\sqrt{T}\mathbf{X_T}(\mathbf{r}) = \mathbf{T}^{-\frac{\mathbf{1}}{\mathbf{2}}}(\boldsymbol{\Psi}(\mathbf{1}) \sum_{s=1}^{t} \boldsymbol{\varepsilon_s} + \boldsymbol{\eta}_{[\mathbf{Tr}]*} - \boldsymbol{\eta_0})$$

The expression $\boldsymbol{\eta}_{[\mathbf{Tr}]*} - \boldsymbol{\eta_0}$ approximates to zero value for $T \to \infty$. In other words it approximates zero.

$$\sup_{i=1,2,\ldots,n} T^{-1/2} \left|\boldsymbol{\eta}_{\boldsymbol{i},[\mathbf{Tr}]*} - \boldsymbol{\eta}_{\mathbf{i},\mathbf{0}}\right| \xrightarrow{P} 0$$

(Hamilton 1995). At this point, using the previously given asymptotic properties the equation below is written:

$$\sqrt{T}\mathbf{X_T}(.) \xrightarrow{P} \boldsymbol{\Psi}(\mathbf{1})\mathbf{P}.\sqrt{\mathrm{T}}\mathbf{X_T^*}(.) \xrightarrow{\boldsymbol{d}} \boldsymbol{\Psi}(\mathbf{1})\mathbf{P}.\mathbf{W}(.)$$

$\boldsymbol{\Psi}(\mathbf{1})\mathbf{P}.\mathbf{W}(\mathbf{r})$ process has the $\boldsymbol{N}(\mathbf{0}, \boldsymbol{r}.\boldsymbol{\Psi}(\mathbf{1}).\boldsymbol{\Omega}.\boldsymbol{\Psi}(\mathbf{1})')$ distribution.

It was previously mentioned that $\boldsymbol{\Omega}$matrix is decomposed as $\boldsymbol{PP'}$ using the cholesky decomposition.The properties below can be used for investigating the asymptotic properties for all multivariate unit rooted series.

$$\boldsymbol{\Gamma_s} = E(\mathbf{a_t a'_{t-s}}) = \sum_{v=0}^{\infty} \boldsymbol{\Psi_{s+v} \Omega \Psi'}_v \quad s = 0,1,2, \ldots$$

$$\mathbf{z_t} = \begin{bmatrix} \mathbf{a_{t-1}} \\ \mathbf{a_{t-2}} \\ \vdots \\ \mathbf{a_{t-v}} \end{bmatrix} \quad v \geq 1$$

$$\boldsymbol{V} = E(\mathbf{z_t z'_t}) = \begin{bmatrix} \boldsymbol{\Gamma_0} & \boldsymbol{\Gamma_1} & \ldots & \boldsymbol{\Gamma_{v-1}} \\ \boldsymbol{\Gamma_{-1}} & \boldsymbol{\Gamma_0} & \ldots & \boldsymbol{\Gamma_v} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\Gamma_{-v+1}} & \boldsymbol{\Gamma_{-v+2}} & \ldots & \boldsymbol{\Gamma_0} \end{bmatrix}$$

$$\boldsymbol{\Lambda} = \boldsymbol{\Psi}(\mathbf{1})\mathbf{P} = (\boldsymbol{\Psi_0} + \boldsymbol{\Psi_1} + \boldsymbol{\Psi_2} + \cdots)\boldsymbol{P}$$

$$\boldsymbol{\Xi_t} = \mathbf{a_1} + \mathbf{a_2} + \cdots + \mathbf{a}_t \,, \boldsymbol{\Xi_0} = \mathbf{0}, \;\; \mathrm{t} = 1,2, \ldots, \mathrm{T}$$

Table 3.2 is an attempt to summarize all asymptotic properties.

Table 3.2 Asymptotic properties for multivariate unit rooted time series

$T^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{a_t} \xrightarrow{L} \boldsymbol{\Lambda W(1)}$;
$T^{-\frac{1}{2}}\sum_{\mathbf{t=1}}^{\mathbf{T}}\mathbf{z_t}\,\boldsymbol{\varepsilon}_{\mathbf{it}} \xrightarrow{L} N(\mathbf{0},\sigma_{ii}\mathbf{V}) \quad i = 1,2,\dots,n$
$T^{-1}\sum_{\mathbf{t=1}}^{\mathbf{T}}\mathbf{a_t}\,\mathbf{a'}_{\mathbf{t-s}} \xrightarrow{p} \boldsymbol{\Gamma_s}$
$T^{-1}\sum_{t=1}^{T}(\boldsymbol{\Xi}_{\mathbf{t-1}}\mathbf{a'}_{\mathbf{t-s}} + \mathbf{a}_{\mathbf{t-s}}\boldsymbol{\Xi'}_{\mathbf{t-1}}) \xrightarrow{d} \begin{cases} \boldsymbol{\Lambda W(1)W'(1)\Lambda' - \Gamma_0} & s = 0 \\ \boldsymbol{\Lambda W(1)W'(1)\Lambda' + \sum_{v=-s+1}^{s-1}\Gamma_v} & s = 1,2,\dots \end{cases}$
$T^{-1}\sum_{\mathbf{t=1}}^{\mathbf{T}}\boldsymbol{\Xi}_{\mathbf{t-1}}\,\mathbf{a'}_{\mathbf{t}} \xrightarrow{d} \boldsymbol{\Lambda}\left\{\int_0^1 \boldsymbol{W(r)(dW(r))'}\right\}\boldsymbol{\Lambda'} + \sum_{\boldsymbol{v=1}}^{\infty}\boldsymbol{\Gamma'}_{\mathbf{v}}$
$T^{-1}\sum_{\mathbf{t=1}}^{\mathbf{T}}\boldsymbol{\Xi}_{\mathbf{t-1}}\,\boldsymbol{\varepsilon'}_{\mathbf{t}} \xrightarrow{L} \boldsymbol{\Lambda}\left\{\int_0^1 \boldsymbol{W(r)(dW(r))'}\right\}\mathbf{P'}$
$T^{-3/2}\sum_{\mathbf{t=1}}^{\mathbf{T}}\boldsymbol{\Xi}_{\mathbf{t-1}} \xrightarrow{L} \boldsymbol{\Lambda}\int_0^1 \mathbf{W(r)dr}$
$T^{-3/2}\sum_{\mathbf{t=1}}^{\mathbf{T}}\mathbf{ta}_{\mathbf{t-s}} \xrightarrow{L} \boldsymbol{\Lambda}\left\{\mathbf{W(1)} - \int_0^1 \mathbf{W(r)dr}\right\}\mathrm{s} = 0,1,2,\dots$
$T^{-2}\sum_{t=1}^{T}\boldsymbol{\Xi}_{\mathbf{t-1}}\boldsymbol{\Xi'}_{\mathbf{t-1}} \xrightarrow{L} \boldsymbol{\Lambda}\left\{\int_0^1 \mathbf{W(r)W'(r)dr}\right\}\boldsymbol{\Lambda'}$
$T^{-5/2}\sum_{\mathbf{t=1}}^{\mathbf{T}}\mathbf{t}\boldsymbol{\Xi}_{\mathbf{t-1}} \xrightarrow{L} \boldsymbol{\Lambda}\int_0^1 \mathbf{rW(r)dr}$
$T^{-3}\sum_{t=1}^{T}\mathbf{t}\boldsymbol{\Xi}_{\mathbf{t-1}}\boldsymbol{\Xi'}_{\mathbf{t-1}} \xrightarrow{L} \boldsymbol{\Lambda}\left\{\int_0^1 \mathbf{rW(r)W'(r)dr}\right\}\boldsymbol{\Lambda'}$

CHAPTER FOUR

COINTEGRATED VAR (1) PROCESSES

The first definition of cointegration was put forth by Engle and Granger (1987) with a theory which argued that the linear combinations of non-stationary series were stationary. They defined the stationary combination of these non-stationary series as cointegration and showed this with the CI (d,b) notation. Here d indicates the integration level of the non-stationary processes, and b indicates the number of linear cointegrations between the non-stationary processes. If (x_{1t} ve x_{2t})the linear combination of two different I(1) series, is stationary [I(0)], then it is said that these two I(1) series are cointegrated and the CI(1,1) notation is used.

Engle and Granger give the definitions below for a CI (d,b) $X_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ vector process with n variables:

a) All n components of the $\mathbf{X_t}$ vector process are cointegrated at d^{th} level.

b) For the β vector, different from zero, $\beta' X_t \sim I(d-b)$ is stationary and β vector is the cointegration vector.

$$\Delta \mathbf{X_t} = \mathbf{\Pi X_{t-1}} + \mathbf{a_t} \tag{4.1}$$

In equation 4.1, $\mathbf{\Pi}$ matrix has a 0<r<n dimensional reduced rank, herefore it can be written as $\mathbf{\Pi = \alpha \beta'}$. Here $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are (nxr) dimensional and have a rank of r. β is the cointegration matrix, α is the adjustment coefficient (loading)matrix, and $\boldsymbol{\alpha'_{\perp}} \mathbf{X_t}$ is a unit rooted process. If r=0 then $\Delta \mathbf{X_t}$ is a stationary VAR process, and if r=n then $\mathbf{X_t}$ is a stationary process.

4.1 Bivariate Cointegrated VAR (1) Process

To be understood more easily, the cointegrated VAR (1) process and the acquisition of cointegration matrix will be explained over a bivariate lateral vector autoregressive process. $\mathbf{X_t} = (X_{1t}\ X_{2t})'$ vector time series is written in terms of a vector autoregressive process in Equation 4.2.

$$\mathbf{X_t} = \mathbf{A X_{t-1}} + \mathbf{a_t} \tag{4.2}$$

The roots of the characteristics equation of the coefficient matrix **A**, under the assumption that the one of the two roots of 4.2 model would have unit root and the other one would be smaller than one, will be as follows:

$$|\mathbf{A} - \mathbf{zI_2}| = 0$$

$$z_1 = 1 \qquad z_2 = \lambda < 1$$

Since $z_1 = 1$, $X_{1t}$and X_{2t} is first order cointegrated and A matrix has a full rank. Therefore A matrix is expressed as eigenvalues and eigenvectors corresponding these eigenvalues.

$$\mathbf{A} = P\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}Q$$

P matrix is the eigenvector matrix of **A**.

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad Q = P^{-1}$$

For ease, under the assumption that the determinant value of P matrix is equal to one, the elements of **A** matrix are expressed as below in terms of their own eigenvalues and eigenvectors.

$$A = \begin{bmatrix} (ad - \lambda bc) & -ab(1-\lambda) \\ cd(1-\lambda) & (-cb + \lambda ad) \end{bmatrix}$$

When both sides of Equation 4.2 are multiplied with Q matrix, being $\begin{bmatrix} w_t \\ -z_t \end{bmatrix} = Q\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}$, the equation below is obtained:

$$\begin{bmatrix} w_t \\ -z_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}\begin{bmatrix} w_{t-1} \\ -z_{t-1} \end{bmatrix} + e_t \tag{4.3}$$

w_t and z_t variables can be written in terms of X_{1t} and X_{2t} variables as below:

$$w_t = dX_{1t} - bX_{2t}$$

$$-z_t = cX_{1t} - aX_{2t}$$

When the equation above is examined, a non-stationary variable such as z_t can be obtained from the linear combination of the non-stationary variables X_{1t} and X_{2t}. In this case, X_{1t} and X_{2t} are cointegrated and $[\mathrm{c} \ -\mathrm{a}]'$ is the cointegration vector.

Another important issue about cointegration is that the cointegration vector is not unique. As it is previously mentioned, the error correction model coefficient matrix in a cointegrated model could be written as $\boldsymbol{\Pi = \alpha\beta'}$. Using the (rxr) dimensional nonsingular C matrix, $\boldsymbol{\alpha^* = \alpha C'}$ and $\boldsymbol{\beta^* = \beta C^{-1}}$ are obtained and this disintegration shows that the cointegration matrix is not unique. This issue can be removed by limiting the cointegration matrix appropriately. $\boldsymbol{\beta}$ matrix has a rank of r; therefore it has r rows linearly independent from each other. Organizing the variables in the model appropriate, and using the information that the first r rows of the cointegration matrix are independent from each other, the cointegration matrix can be made unique. For this purpose, the cointegration vector can be selected as below:

$$\boldsymbol{\beta^*} = \begin{bmatrix} \mathbf{I_r} \\ \boldsymbol{\beta_{K-r}} \end{bmatrix} \tag{4.4}$$

$\boldsymbol{\beta_{K-r}}$ is the (n-r)xr dimensional matrix, and $\mathbf{I_r}$ is the unit matrix. Organization of $\boldsymbol{\beta^*}$ as in Equation 4.4 is called normalization. Using this normalization, the cointegration vector is made unique. The results of the normalization operation will be explained over a trivariate system with a rank of 1 and with all variables first order cointegrated (I(1)). In this system the model that shows cointegration can be written as $\boldsymbol{\beta^* Y_t} = [1 \ , \boldsymbol{\beta'_{K-r}}]\mathbf{Y_t} = \mathrm{y_{1t}} + \beta_2 \mathrm{y_{2t}} + \beta_3 \mathrm{y_{3t}}$. In order for this normalization to apply, $\mathrm{y_{1t}}$ needs to be integrated into the cointegration relationship and its coefficient is different from zero. Although $\mathrm{y_{2t}}$ and $\mathrm{y_{3t}}$ are not cointegrated, the condition that the variables $\mathrm{y_{1t}}$, $\mathrm{y_{2t}}$ and $\mathrm{y_{3t}}$ are cointegrated together leads to the result that $\mathrm{y_{1t}}$ is naturally integrated into the cointegration relationship; and therefore its coeficient in $\boldsymbol{\beta}$ vector is different from zero.

4.2 Estimation of Bivariate Cointegrated VAR (1) Process

Consider bivariate cointegrated VAR (1) process is as follows:

$$\Delta \boldsymbol{X_t} = \mathbf{\Pi X_{t-1}} + \boldsymbol{a_t} = \boldsymbol{\alpha\beta'}\mathbf{X_{t-1}} + \boldsymbol{a_t} \quad t = 1,2,\ldots \tag{4.5}$$

where $\mathbf{\Pi}$ is (2x2) matrix of rank r=1 (0<r<2), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are (2x1) with rank r=1 and $\mathbf{a_t}$ is two dimensional white noise process with mean zero and variance-covariance matrix $\mathbf{\Sigma_a}$. Also we suppose that $\mathbf{X_t}$ is I (1) process and $\boldsymbol{\alpha'_\perp\beta_\perp}$ is an invertible, because it is real valued scalar. $\boldsymbol{\beta_\perp}$ and $\boldsymbol{\alpha_\perp}$ are orthogonal complements of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. If r=0, then $\Delta\mathbf{X_t}$ is stationary and if r=p=2 then $\mathbf{X_t}$ is stationary.

Maximum Likelihood and Unrestricted Least Square estimator of $\mathbf{\Pi}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are discussed in this section. Then asymptotic distribution of these related estimators are derived.

Unrestricted LS estimation method is preferred to LS estimation method due to the lack of the variance information. Using normal equations, unrestricted LS estimator of $\mathbf{\Pi}$ is obtained as follows:

$$\widehat{\mathbf{\Pi}} = \left(\sum_{t=1}^{T} \Delta \boldsymbol{X_t X'_{t-1}}\right)\left(\sum_{t=1}^{T} \boldsymbol{X_{t-1} X'_{t-1}}\right)^{-1} \tag{4.6}$$

if $\mathbf{\Pi X_{t-1}} + \boldsymbol{a_t}$ is replaced instead of $\Delta \boldsymbol{X_t}$, then equation 4.7. is obtained.

$$\widehat{\mathbf{\Pi}} - \mathbf{\Pi} = \left(\sum_{t=1}^{T} \boldsymbol{a_t X'_{t-1}}\right)\left(\sum_{t=1}^{T} \boldsymbol{X_{t-1} X'_{t-1}}\right)^{-1} \tag{4.7}$$

$\mathbf{Q}$ (2x2) matrix can be choosen as follows,

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{\beta'} \\ \boldsymbol{\alpha'_\perp} \end{bmatrix} \quad , \quad \boldsymbol{Q^{-1}} = [\boldsymbol{\alpha}(\boldsymbol{\beta'\alpha})^{-1} \quad \boldsymbol{\beta_\perp}(\boldsymbol{\alpha_\perp}'\boldsymbol{\beta_\perp})^{-1}]$$

If the left hand side of equation of 4.7 is multiplied by $\boldsymbol{Q}$ and the right hand side of the equation is multiplied by $\boldsymbol{Q^{-1}}$, following equation can be obtained

$$\boldsymbol{Q}\left(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}\right)\boldsymbol{Q}^{-1}=\boldsymbol{Q}\left(\sum_{t=1}^{T}\boldsymbol{a}\boldsymbol{X}_{t-1}'\right)\boldsymbol{Q}'\boldsymbol{Q}^{-1'}\left(\sum_{t=1}^{T}\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1}\boldsymbol{Q}^{-1}$$

$$=\left(\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{z}_{t-1}'\right)\left(\sum_{t=1}^{T}\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}'\right)^{-1}$$

where $\boldsymbol{v_t}=\boldsymbol{Q}\boldsymbol{a_t}$ and $\boldsymbol{z_t}=\boldsymbol{Q}\boldsymbol{X_t}$.

Hence, denoting the first r components of $\boldsymbol{z_t}$ by $\boldsymbol{z_t}^{(1)}=\boldsymbol{\beta}'\mathbf{X_t}$ which consists of the cointegration relationship and therefore the stationarity while the last K-r components of $\boldsymbol{z_t}$, denoted by $\boldsymbol{z_t}^{(2)}=\boldsymbol{\alpha}_{\perp}{}'\mathbf{X_t}$ which contains a K-r dimensional random walk because $\Delta\boldsymbol{z_t}^{(2)}$ is white noise. So, z_t is separated into two parts - former is stationary and latter is nonstationary.

To derive the asymptotic properties of the LS estimator, it is useful to rewrite

$$\boldsymbol{Q}\left(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}\right)\boldsymbol{Q}^{-1}$$

$$=\left[\sum_{t=1}^{T}\boldsymbol{v}_t\,\boldsymbol{z}_{t-1}{}^{(1)\prime}\quad\sum_{t=1}^{T}\boldsymbol{v}_t\,\boldsymbol{z}_{t-1}{}^{(2)\prime}\right]\begin{bmatrix}\sum_{t=1}^{T}\boldsymbol{z}_{t-1}{}^{(1)}\,\boldsymbol{z}_{t-1}{}^{(1)\prime} & \sum_{t=1}^{T}\boldsymbol{z}_{t-1}{}^{(1)}\,\boldsymbol{z}_{t-1}{}^{(2)\prime}\\ \sum_{t=1}^{T}\boldsymbol{z}_{t-1}{}^{(2)}\,\boldsymbol{z}_{t-1}{}^{(1)\prime} & \sum_{t=1}^{T}\boldsymbol{z}_{t-1}{}^{(2)}\,\boldsymbol{z}_{t-1}{}^{(2)\prime}\end{bmatrix}^{-1}$$

Table 4.1 Assymptotic properties of stationary and nonstationary process

1	$\boldsymbol{T^{-1}\sum_{t=1}^{T} z_{t-1}{}^{(1)}\ z_{t-1}{}^{(1)\prime} \xrightarrow{P} \Gamma_z^{(1)}}$. where $\Gamma_z^{(1)}$ is the covariance matrix of $z_t{}^{(1)}$
2	$\boldsymbol{T^{-\frac{1}{2}} vec\left(\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(1)\prime}\right) \xrightarrow{d} N\left(0, \Gamma_z^{(1)} \otimes \Sigma_v\right)}$
3	$\boldsymbol{T^{-1}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(2)\prime} \xrightarrow{d} \Sigma_v^{1/2}\left(\int_0^1 W_k dW_k'\right)' \Sigma_v^{1/2} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}}$, Where $\boldsymbol{W_K}$ denotes the standard wiener process $\boldsymbol{W_K(s)}$ of dimension K.
4	$\boldsymbol{T^{-3/2} \sum_{t=1}^{T} z_{t-1}{}^{(1)}\ z_{t-1}{}^{(2)\prime} \xrightarrow{p} 0}$.
5	$\mathbf{T^{-2} \sum_{t=1}^{T} z_{t-1}{}^{(2)}\ z_{t-1}{}^{(2)\prime} \xrightarrow{d} [0 \quad I_{K-r}]\ \Sigma_v^{\frac{1}{2}} \left(\int_0^1 W_k W_k' ds\right) \Sigma_v^{\frac{1}{2}} \begin{bmatrix} 0 \\ I_{K-r} \end{bmatrix}}$

Ahn & Reinsel (1990) would be helpful for details in derivation of the asymptotic distribution of $\boldsymbol{\hat{\Pi} - \Pi}$. The following information in table 4.1 will use frequently in the other sections.

4.2.1 Limiting Results for the LS Estimator $\hat{\Pi}$

D matrix is considered as follows:

$$\boldsymbol{D} = \begin{bmatrix} T^{1/2} & 0 \\ 0 & T \end{bmatrix}$$

where its elements, $T^{1/2}$ and T, are convergence rates.

Then

$$\boldsymbol{vec}[\boldsymbol{Q}(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi})\boldsymbol{Q}^{-1}\boldsymbol{D}]$$

$$\xrightarrow{d}\begin{bmatrix} N(\mathbf{0},(\Gamma_z^{(1)})^{-1}\otimes\sum_{\mathrm{v}}) \\ vec\left\{\Sigma_v^{\frac{1}{2}}(\int_0^1 W_k W_k' ds)'\Sigma_v^{\frac{1}{2}}\begin{bmatrix}\mathbf{0}\\ I_{K-r}\end{bmatrix}([\mathbf{0}\quad I_{K-r}]\Sigma_v^{\frac{1}{2}}(\int_0^1 W_k W_k' ds)\Sigma_v^{\frac{1}{2}}\begin{bmatrix}\mathbf{0}\\ I_{K-r}\end{bmatrix})^{-1}\right\} \end{bmatrix}$$

The $vec[\boldsymbol{Q}(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi})\boldsymbol{Q}^{-1}\boldsymbol{D}]$ is distributed as a combination of normal distribution and Wiener process.

Proof:

$$\boldsymbol{Q}(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi})\boldsymbol{Q}^{-1}\boldsymbol{D}$$

$$=\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(1)\prime} & T^{-1}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(2)\prime}\end{bmatrix} D \begin{bmatrix} \sum_{t=1}^{T} z_{t-1}{}^{(1)} z_{t-1}{}^{(1)\prime} & \sum_{t=1}^{T} z_{t-1}{}^{(1)} z_{t-1}{}^{(2)\prime} \\ \sum_{t=1}^{T} z_{t-1}{}^{(2)} z_{t-1}{}^{(1)\prime} & \sum_{t=1}^{T} z_{t-1}{}^{(2)} z_{t-1}{}^{(2)\prime}\end{bmatrix}^{-1} D$$

=

$$\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(1)\prime} & T^{-1}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(2)\prime}\end{bmatrix} \begin{bmatrix} T^{-1}\sum_{t=1}^{T} z_{t-1}{}^{(1)} z_{t-1}{}^{(1)\prime} & T^{-1}\sum_{t=1}^{T} z_{t-1}{}^{(1)} z_{t-1}{}^{(2)\prime} \\ T^{-2}\sum_{t=1}^{T} z_{t-1}{}^{(2)} z_{t-1}{}^{(1)\prime} & T^{-2}\sum_{t=1}^{T} z_{t-1}{}^{(2)} z_{t-1}{}^{(2)\prime}\end{bmatrix}^{-1}$$

Using by partitioned inverse, following matrix is yield as follows:

$$=\begin{bmatrix} T^{-1/2}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(1)\prime} & T^{-1}\sum_{t=1}^{T} v_t\, z_{t-1}{}^{(2)\prime}\end{bmatrix} \begin{bmatrix} S_{11}^{-1}+S_{11}^{-1}S_{12}S^{*}S_{21}S_{11}^{-1} & -S_{11}^{-1}S_{12}S^{*} \\ -S^{*}S_{21}S_{11}^{-1} & S^{*}\end{bmatrix}$$

where $\boldsymbol{S^{*}=(S_{22}^{-1}-S_{21}S_{11}^{-1}S_{12})^{-1}}$.

By using first information in table4.1,

$$\boldsymbol{S_{11}} = T^{-1}\sum_{t=1}^{T} z_{t-1}{}^{(1)} z_{t-1}{}^{(1)\prime} \xrightarrow{p} \boldsymbol{\Gamma_z^{(1)}}$$

the $\boldsymbol{S_{11}}$ is converging in probabilityto variance-covariance matrix $\Gamma_z^{(1)}$ of stationary process $(\boldsymbol{z_t}^{(1)} = \boldsymbol{\beta}'\mathbf{X_t})$.

By using 4th information in table 4.1,

$$\boldsymbol{S_{12}} = \boldsymbol{S_{21}'} = T^{-1}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{z_{t-1}}^{(1)}\, \boldsymbol{z_{t-1}}^{(2)\prime} \xrightarrow{d} \boldsymbol{o_p(T^{1/2})}$$

the $\boldsymbol{S_{12}}$ is converging in distribution to zero with converging rate $T^{(-3/2)}$.

By using 5th information of Table 4.1 and the continuous mapping theorem;

$$\boldsymbol{S_{22}} = T^{-2}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{z_{t-1}}^{(2)}\, \boldsymbol{z_{t-1}}^{(2)\prime} = \boldsymbol{O_p(1)}$$

$$\boldsymbol{S_{22}^{-1}} = \boldsymbol{O_p(1)}$$

The inverse of $\boldsymbol{S_{22}}$ convergences to a real-valued scalar $[\mathbf{0} \quad \boldsymbol{I_{K-r}}]\, \boldsymbol{\Sigma_v^{\frac{1}{2}}}(\int_0^1 \boldsymbol{W_k W_k' ds})\boldsymbol{\Sigma_v^{\frac{1}{2}}} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{I_{K-r}} \end{bmatrix}$ with convergence rate $T^{(-2)}$.

Using rules of partitioned inverse;

$$\boldsymbol{S^*} = \boldsymbol{S_{22}^{-1}} + \boldsymbol{S_{22}^{-1}S_{21}}\left(\boldsymbol{S_{11}} - \boldsymbol{S_{12}S_{22}^{-1}S_{21}}\right)^{-1}\boldsymbol{S_{12}S_{22}^{-1}}$$

$$= \boldsymbol{O_p(1)} + \boldsymbol{O_p(1)O_p}\left(\boldsymbol{T^{-\frac{1}{2}}}\right)\boldsymbol{O_p(1)o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)\boldsymbol{O_p(1)} = \boldsymbol{O_p(1)}$$

S* convergences to finite real-valued scalar since $\boldsymbol{o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)$ convergences to zero.

It can be seen easily, $\boldsymbol{S_{11} - S_{12}S_{22}^{-1}S_{21}}$ convergences to a real-valued scalar.

$$\boldsymbol{S_{11} - S_{12}S_{22}^{-1}S_{21}} = \boldsymbol{S_{11}} - \boldsymbol{o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)\boldsymbol{O_p(1)o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right) = \boldsymbol{S_{11}}$$

Based on continuous mapping theorem, the inverse of $\boldsymbol{S_{11} - S_{12}S_{22}^{-1}S_{21}}$ also convergences to the scalar.

$$\left(\boldsymbol{S_{11} - S_{12}S_{22}^{-1}S_{21}}\right)^{-1} = \boldsymbol{O_p(1)}$$

As a result,

$$\boldsymbol{S_{11}^{-1} + S_{11}^{-1}S_{12}S^{*}S_{21}S_{11}^{-1}} = \left(\boldsymbol{\Gamma_z^{(1)}}\right)^{-1} + \boldsymbol{O_p(1)o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)\boldsymbol{O_p(1)o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)\boldsymbol{O_p(1)}$$

$$= (\boldsymbol{\Gamma_z^{(1)}})^{-1} + \boldsymbol{o_p(1)}$$

and

$$-\boldsymbol{S_{11}^{-1}S_{12}S^{*}} = -\boldsymbol{O_p(1)o_p}\left(\boldsymbol{T^{\frac{1}{2}}}\right)\boldsymbol{O_p(1)} = \boldsymbol{o_p(1)}$$

Thus,

$$= \left[T^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(1)\prime} \quad \boldsymbol{T^{-1}}\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(2)\prime}\right]$$

$$\times\begin{bmatrix}(\boldsymbol{T^{-1}}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(1)}\,\boldsymbol{z_{t-1}}^{(1)\prime})^{-1} + \boldsymbol{o_p(1)} & \boldsymbol{o_p(1)} \\ \boldsymbol{o_p(1)} & (\boldsymbol{T^{-2}}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(2)}\,\boldsymbol{z_{t-1}}^{(2)\prime})^{-1} + \boldsymbol{o_p(1)}\end{bmatrix}$$

$$=[\ T^{-\frac{1}{2}}\textstyle\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(1)\prime}(T^{-1}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(1)}\,\boldsymbol{z_{t-1}}^{(1)\prime})^{-1}$$

$$T^{-1}\textstyle\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(2)\prime}(T^{-2}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(2)}\,\boldsymbol{z_{t-1}}^{(2)\prime})^{-1}]+o_p(1)$$

Finally,

$$vec\left[\boldsymbol{Q}\left(\hat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)\boldsymbol{Q^{-1}D}\right]$$

$$= \begin{bmatrix} vec(T^{-\frac{1}{2}}\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(1)\prime}(T^{-1}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(1)}\,\boldsymbol{z_{t-1}}^{(1)\prime})^{-1}) \\ vec(T^{-1}\sum_{t=1}^{T}\boldsymbol{v_t}\,\boldsymbol{z_{t-1}}^{(2)\prime}\left(T^{-2}\sum_{t=1}^{T}\boldsymbol{z_{t-1}}^{(2)}\,\boldsymbol{z_{t-1}}^{(2)\prime}\right)^{-1}) \end{bmatrix}$$

Using table 4.1, the proof has been completed.

$$\xrightarrow{d}\begin{bmatrix} N(\mathbf{0},(\boldsymbol{\Gamma}_z^{(1)})^{-1}\otimes\sum_{\mathbf{v}} \\ vec\left\{\boldsymbol{\Sigma}_v^{\frac{1}{2}}(\int_0^1 \boldsymbol{W_k W_k' ds})'\boldsymbol{\Sigma}_v^{\frac{1}{2}}\begin{bmatrix}\mathbf{0}\\ \boldsymbol{I_{K-r}}\end{bmatrix}([\mathbf{0}\quad \boldsymbol{I_{K-r}}]\boldsymbol{\Sigma}_v^{\frac{1}{2}}(\int_0^1 \boldsymbol{W_k W_k' ds})\boldsymbol{\Sigma}_v^{\frac{1}{2}}\begin{bmatrix}\mathbf{0}\\ \boldsymbol{I_{K-r}}\end{bmatrix})^{-1}\right\} \end{bmatrix}$$

The $vec\left[\boldsymbol{Q}\left(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi}\right)\boldsymbol{Q^{-1}D}\right]$ is still consisting of nonnormal elements. Choosing proper convergence rate, the nonnormal part of matrix could be normal.

The distribution of unrestricted LSE estimator $\widehat{\boldsymbol{\Pi}}$ is asymptotically normal,

$$\sqrt{T}vec(\widehat{\boldsymbol{\Pi}}-\boldsymbol{\Pi})\xrightarrow{d}N(\mathbf{0},\boldsymbol{\beta}\left(\boldsymbol{\Gamma}_z^{(1)}\right)^{-1}\boldsymbol{\beta}'\otimes\boldsymbol{\Sigma_u})$$

And $\boldsymbol{\beta}\left(\boldsymbol{\Gamma}_z^{(1)}\right)^{-1}\boldsymbol{\beta}'$ is estimated by using $(T^{-1}\sum_{t=1}^{T}\boldsymbol{X_{t-1}X_{t-1}'})^{-1}$

4.2.2 Limiting Results for the MLE Estimator $\widehat{\Pi}$

When the error process is assumed to be Normal distribution, maximum likelihood estimator can be used to estimate unknown parameters. If $\boldsymbol{\alpha}$ and $\sum_a$ are known, the maximum likelihood estimator is the same as Generalized Least Sqaure (GLS) estimator for $\widehat{\boldsymbol{\beta}}'_{\boldsymbol{K-r}}$. The log likelihood function is given as following:

$$\ln(l)=-\frac{KT}{2}ln2\pi-\frac{T}{2}ln|\boldsymbol{\Sigma_a}|-\frac{1}{2}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}(\Delta\boldsymbol{y_t}-\boldsymbol{\Pi X_{t-1}})'\boldsymbol{\Sigma_u^{-1}}(\Delta\boldsymbol{X_t}-\boldsymbol{\Pi X_{t-1}})$$

Maximizing log-likelihood function is possible just minimizing the following determinant.

$$\left|T^{-1}\sum_{t=1}^{T}(\Delta\boldsymbol{y_t}-\boldsymbol{\Pi X_{t-1}})(\Delta\boldsymbol{y_t}-\boldsymbol{\Pi X_{t-1}})'\right|$$

For the general case, rank ($\boldsymbol{\Pi}$) =r, it means that there are r cointegration relationship. We can write $\boldsymbol{\Pi}=\boldsymbol{\alpha\beta}'$, so the determinant is given by

$$\left| T^{-1} \sum_{t=1}^{T} (\Delta \boldsymbol{X_t} - \boldsymbol{\alpha\beta'} \boldsymbol{X_{t-1}})(\Delta \boldsymbol{X_t} - \boldsymbol{\alpha\beta'} \boldsymbol{X_{t-1}})' \right|$$

with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The minimum value of the determinant is attained for

$$\widetilde{\boldsymbol{\beta}} = [v_1 \quad , \dots , \quad v_r]' (\sum_{t=1}^{T} \boldsymbol{X_{t-1}} \boldsymbol{X'_{t-1}})^{-1/2}$$

$$\widetilde{\boldsymbol{\alpha}} = \left(\sum_{t=1}^{T} \Delta \boldsymbol{X_t} \boldsymbol{X'_{t-1}} \widetilde{\boldsymbol{\beta}}\right)\left(\sum_{t=1}^{T} \widetilde{\boldsymbol{\beta}}' \boldsymbol{X_{t-1}} \boldsymbol{X'_{t-1}} \widetilde{\boldsymbol{\beta}}\right)^{-1}.$$

where the eigenvalues $\lambda_1, \geq \lambda_2 \geq \cdots \geq \lambda_K$ and the associated orthonormal eigenvectors $\boldsymbol{v_1} \quad , \dots , \quad \boldsymbol{v_r}$ is obtained from the following matrix

$$\left(\sum_{t=1}^{T} \boldsymbol{X_t} \boldsymbol{X'_{t-1}}\right)^{-1/2} \left(\sum_{t=1}^{T} \boldsymbol{X_{t-1}} \Delta \boldsymbol{X'_t}\right)\left(\sum_{t=1}^{T} \boldsymbol{X_t} \boldsymbol{X'_t}\right)\left(\sum_{t=1}^{T} \Delta \boldsymbol{X_t} \boldsymbol{X'_{t-1}}\right)\left(\sum_{t=1}^{T} \boldsymbol{X_t} \boldsymbol{X'_{t-1}}\right)^{-1/2}$$

And also $\widetilde{\boldsymbol{\Pi}} = \widetilde{\boldsymbol{\alpha}}\widetilde{\boldsymbol{\beta}}'$ must have same asymptotic results as the unrestricted LS estimator of $\boldsymbol{\Pi}$. We know that $\widehat{\boldsymbol{\beta}}'$ does not affect the LS estimator $\boldsymbol{\Pi}$. And also, MLE estimator of α is equal to LS estimator (Lutkepohl 2005). That is given in the following asymptotic results,

$$\sqrt{T} vec(\widetilde{\boldsymbol{\alpha}}\widetilde{\boldsymbol{\beta}}' - \boldsymbol{\Pi}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\beta}\left(\boldsymbol{\Gamma}_z^{(1)}\right)^{-1} \boldsymbol{\beta}' \otimes \boldsymbol{\Sigma_u})$$

To reach unique $\widehat{\boldsymbol{\beta}}'$, normalized MLE estimator of $\boldsymbol{\beta}$ should be obtained. $\breve{\boldsymbol{\beta}} = \begin{bmatrix} \mathbf{I_r} \\ \breve{\boldsymbol{\beta}}_{\mathbf{K-r}} \end{bmatrix}$ is normalized MLE estimator $\boldsymbol{\beta}$ and also the normalized estimator for MLE estimator $\widetilde{\boldsymbol{\alpha}}$ can be obtained explicitly. $\breve{\boldsymbol{\beta}}$ and $\breve{\boldsymbol{\alpha}}$ estimators are given below:

$$\breve{\alpha} = \left(\sum_{t=1}^{T} \Delta \boldsymbol{X_t} \boldsymbol{X'_{t-1}} \breve{\boldsymbol{\beta}}\right)\left(\sum_{t=1}^{T} \breve{\boldsymbol{\beta}}' \boldsymbol{X_{t-1}} \boldsymbol{X'_{t-1}} \breve{\boldsymbol{\beta}}\right)^{-1}$$

$$\breve{\boldsymbol{\beta}}'_{K-r} = \left(\breve{\boldsymbol{\alpha}}' {\breve{\boldsymbol{\Sigma}}_u}^{-1} \breve{\boldsymbol{\alpha}}\right)^{-1} \breve{\boldsymbol{\alpha}}' {\breve{\boldsymbol{\Sigma}}_a}^{-1} \left(\sum_{t=1}^{T} \left(\Delta \boldsymbol{y_t} - \breve{\boldsymbol{\alpha}} \boldsymbol{X_{t-1}^{(1)}}\right) \boldsymbol{X_{t-1}^{(2)\prime}}\right)\left(\left(\sum_{t=1}^{T} \boldsymbol{X_{t-1}^{(2)}} \boldsymbol{X_{t-1}^{(2)\prime}}\right)^{-1}\right)$$

MLE estimators of $\breve{\boldsymbol{\Pi}}$, $\breve{\boldsymbol{\alpha}}$ and $\breve{\boldsymbol{\beta}}$ have same asymptotic properties as LS estimators of $\hat{\boldsymbol{\Pi}}$, $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. So, asymptotic properties are identical for both estimation techniques.

4.3 Simulation Study for Estimation Under the Unit Root

In this section, finite sample properties of both estimators are considered through Monte Carlo simulation. Cointegrated bivariate model $X_t = AX_{t-1} + u_t$ is simulated with following coefficient matrix,

$$A = \begin{bmatrix} \rho & \theta \\ 0 & \alpha \end{bmatrix},$$

and variance covariance matrix of iid error process

$$\Sigma_a = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Simulation is performed for different ρ and α values in A matrix. Characteristic roots have only one root, either if $\rho = 1$ and $\alpha < 1$ or $\alpha = 1$ and $\rho < 1$. We assume cointegrated process with one unit root.

The aim of the simulation study is to examine the asymptotic properties of $E(\hat{\rho}) - \rho$ and $E(\hat{\alpha}) - \alpha$. Asymptotic properties of these quantities is examined under constant ρ and varying α, then constant α and varying ρ conditions. The important point is that one of these quantities (either α or ρ) should not be greater than one, because we consider one unit root and one stationary root in the bivariate system. In both conditions, θ is the same because its value doesn't affect the stationarity of the system. Then $E(\hat{\rho}) - \rho$ and $E(\hat{\alpha}) - \alpha$ are performed for different replications T=50, 100, 250 through Monte Carlo simulation.

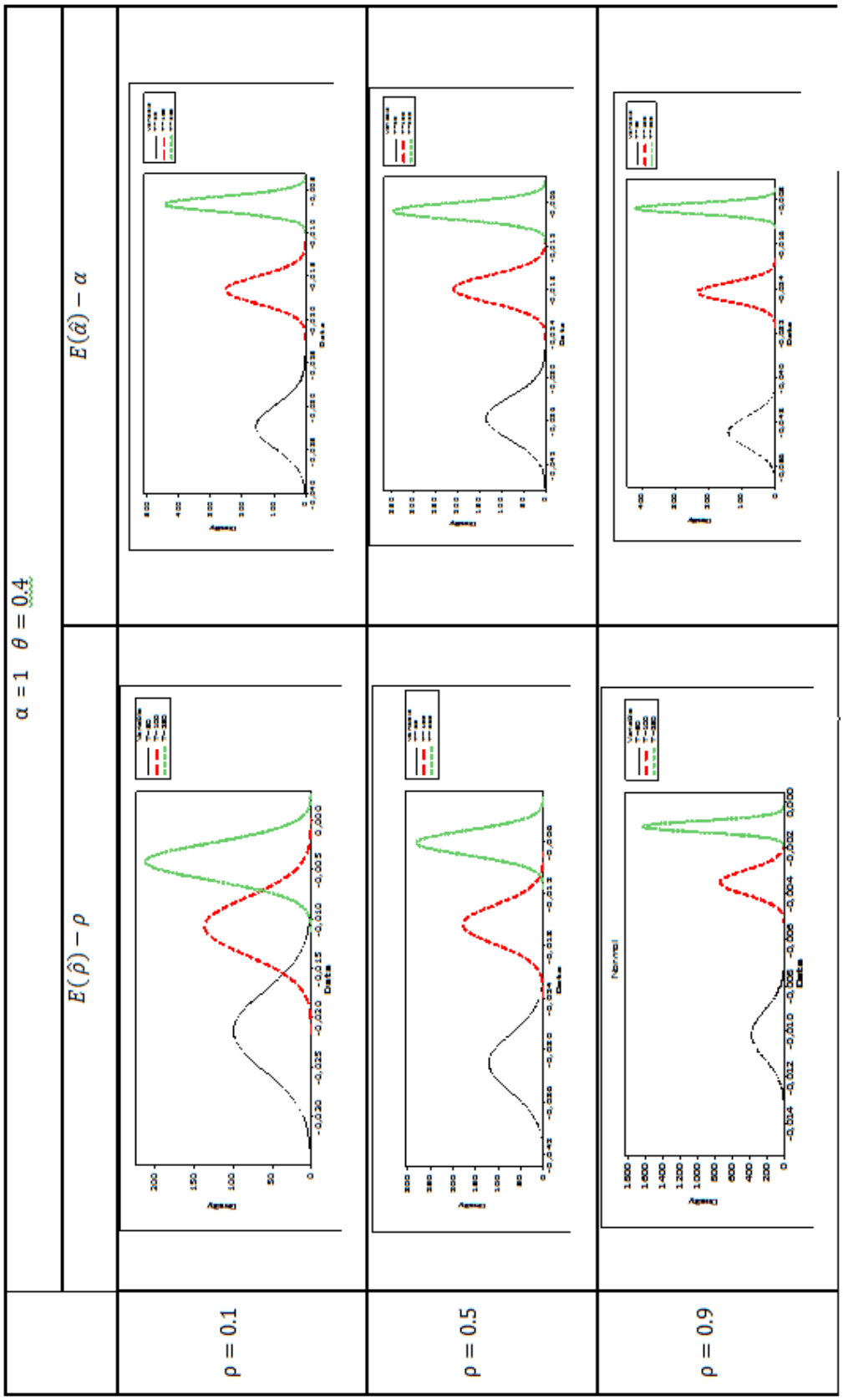

Figure 4.1 Histograms of $E(\hat{\rho}) - \rho$ and $E(\hat{\alpha}) - \alpha$ for α = 1 and ρ =0.1 , 0.5 , 0.9 ; $\theta = 0.4$

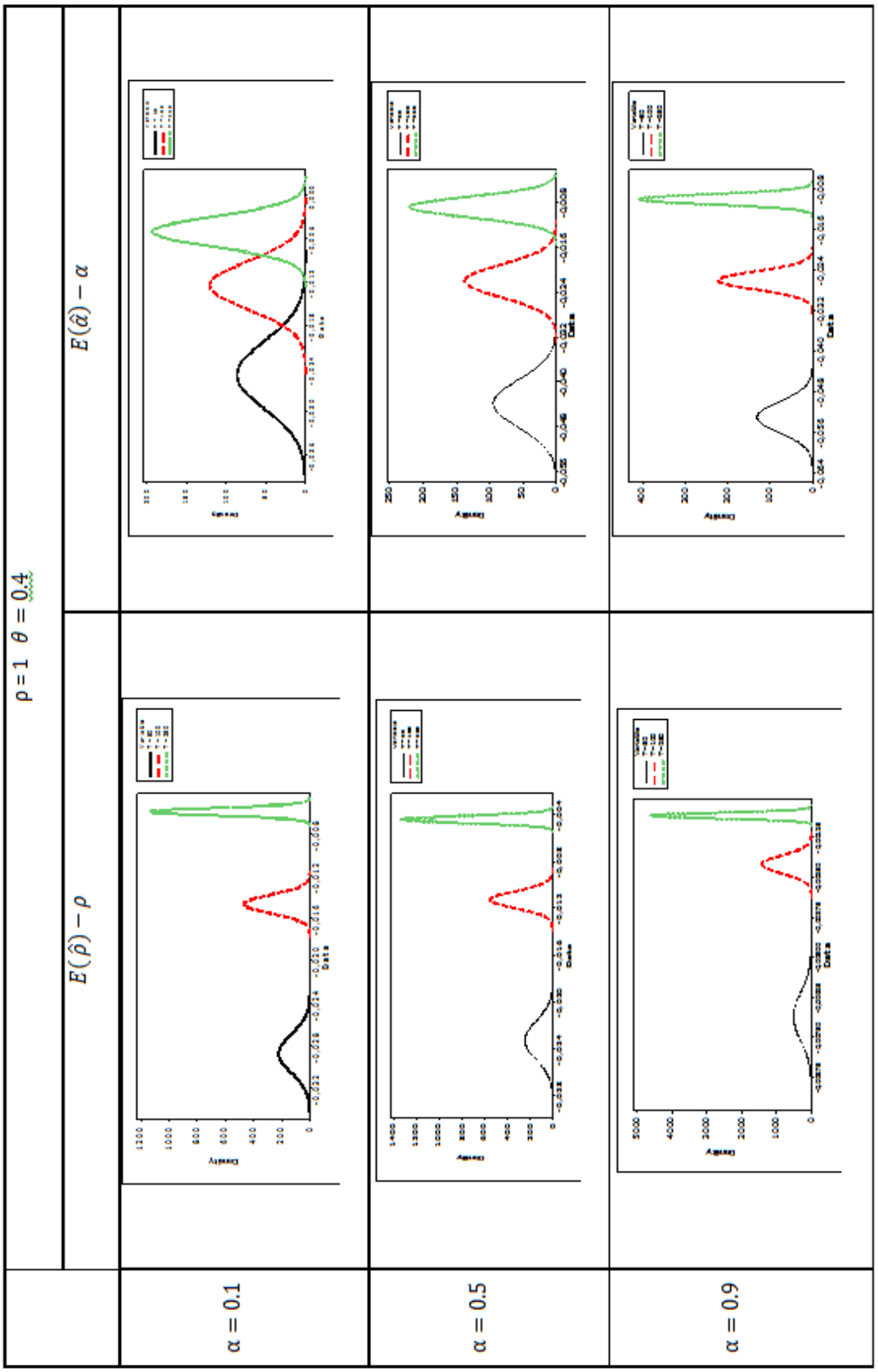

Figure 4.2 Histograms of $E(\hat{\rho}) - \rho$ and $E(\hat{\alpha}) - \alpha$ for $\rho = 1$ and $\alpha = 0.1, 0.5, 0.9$; $\theta = 0.4$

When histograms which are illustrated in Figure 4.1 are examined- $\alpha = 1$ and ρ =0.1 , 0.5 , 0.9- distributions of $E(\hat{\rho}) - \rho$ have smaller variances as T increases for any case in ρ. Also, distributions of $E(\hat{\rho}) - \rho$ have smaller kurtosis and narrower confidence intervals, and less biasness for all ρ. Unlike the distributions of $E(\hat{\rho}) - \rho$, distributions of $E(\hat{\alpha}) - \alpha$ have not changed for all ρ under the same conditions.

As shown in Figure 4.2, for $\rho = 1$, α =0.1, 0.5, 0.9, $\theta = 0.4$, variances of distribution of $E(\hat{\rho}) - \rho$ are increasing considerably in contrast to variances of distribution $E(\hat{\alpha}) - \alpha$ for sample size of 50. Especially distributions of $E(\hat{\alpha}) - \alpha$ have smaller variances for all sample size.

Table 4.2 Mean Square Error of Parameters when $\alpha = 1$

		α =1		
	ρ	**0.1**	**0.5**	**0.9**
T=50	**MSE** $\hat{\alpha}$	0.0011	0.0013	0.0025
	MSE $\hat{\rho}$	0.000474	0.1358	0.6238
T=100	**MSE** $\hat{\alpha}$	0.000277	0.00032	0.00060
	MSE $\hat{\rho}$	0.000125	0.1478	0.6344
T=250	**MSE** $\hat{\alpha}$	0.000045	0.000052	0.000092
	MSE $\hat{\rho}$	0.000021	0.1551	0.6338

As it is shown in Table 4.2; for all cases, as time series length increases, mean square errors (MSE) of $\hat{\alpha}$ and $\hat{\rho}$ parameters decreases. When ρ approaches to one, MSE of parameter $\hat{\rho}$ increases remarkably comparing to α.Reversely, when α approaches to 1, this increasing rate of MSE of $\hat{\alpha}$ and $\hat{\rho}$ parameters is slower than ρ approaches to 1 as shown in Table 4.3.

Table 4.3 Mean Square Error of Parameters when $\rho = 1$

		ρ=1		
	α	**0.1**	**0.5**	**0.9**
T=50	**MSE** $\widehat{\boldsymbol{\alpha}}$	0.000642	0.001938	0.002818
	MSE $\widehat{\boldsymbol{\rho}}$	0.000830	0.000545	0.000047
T=100	**MSE** $\widehat{\boldsymbol{\alpha}}$	0.000167	0.000492	0.000683
	MSE $\widehat{\boldsymbol{\rho}}$	0.000215	0.000127	0.000004
T=250	**MSE** $\widehat{\boldsymbol{\alpha}}$	0.000030	0.000081	0.000102
	MSE $\widehat{\boldsymbol{\rho}}$	0.000035	0.000019	0.000000

When ρ has unit root, the MSE of parameters has better results. In existence of exogenous variables in the bivariate system, the asymptotic properties' of parameters (α and ρ) have better under ρ=1. The properties are almost unbiased and consistent.

CHAPTER FIVE

LIKELIHOOD RATIO TESTS FOR COINTEGRATION RANK AND THE LOCAL POWER OF THESE TESTS

This chapter will discuss the hypothesis tests developed for testing the cointegration rank and the power functions of these tests will be evaluated using local power analysis. Local power analysis is the investigation of the behaviors of power function of the hypothesis tests with neighboring null hypothesis (Macnamus, 1991). As the sample size increases the local alternative hypothesis becomes closer to the null hypothesis and all possible situations are evaluated.

As it was mentioned in the previous chapters, the number of the independent linear cointegration relations of the multivariate time series X_t was being expressed with the rank (r) of the Π matrix in the model below:

$$\Delta X_t = \Pi X_{t-1} + a_t$$

The number of the relations in question can be tested with the hypothesis tests. These hypothesis tests are called as the Trace test by Johansen (1988) and the Maximum EigenValue Test (1995). These two alternative tests differ from each other in terms of the hypotheses formed. The hypothesis for Trace test, varible number being *n*, is as below:

$$H_0(r_0): rk(\Pi) = r_0$$
$$H_1(r_0): rk(\Pi) = n$$

On the other hand the alternative hypothesis in the Maximum EigenValue test is formed as below:

$$H_0(r_0): rk(\Pi) = r_0$$
$$H_1(r_0 + 1): rk(\Pi) = r_0 + 1$$

If hypothesis H_0 is true, then Π matrix is written as $\alpha\beta'$ and hypothesis $H_0(r_0): \Pi = \alpha\beta'$ Here α and β are (nxr_0) dimensional and r_0 ranked matrices. On the other hand, if $rank(\Pi) > r_0$ then Π matrix is expressed as below:

$$\Pi = [\alpha: \alpha_1] \begin{bmatrix} \beta' \\ \beta_1' \end{bmatrix} = \alpha\beta' + \alpha_1\beta_1'$$

The following local alternative hypothesis is formed for local power analysis:

$$H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

By generating simulation data under the assumption that the local hypothesis is true, the asymptotic distribution is obtained for $T \to \infty$.

5.1 Johansen Trace Test and Its Power Under Local Alternative Hypothesis

The model used here does not contain a linear trend without constant terms. This model is as below:

$$Y_t = \Pi X_{t-1} + \varepsilon_t \qquad (5.1)$$

Here the variable matrix X_t, is I(1) first order cointegrated and the Π coefficient matrix of the model has a reduced rank feature. In other words, Π matrix is expressed as a product of β' . In this test the aim is to test the hypotheses below:

$$H_0: r = r_0$$

$$H_1: r = n$$

Since the local power of the Johansen Trace test is at issue here, the local alternative hypothesis is as below:

$$H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

When the model

$$Y_t = AB'X_t + a_t \qquad t = 1, \dots, \mathrm{T} \qquad (5.2)$$

is considered, Y_t and a_t are (nx1) dimensional, m>n, X_t is (mx1) dimensional, A and B matrices are (nxr_0) and (mxr_0) dimensional respectively. If the error term is formed as a form of a_t

$$a_t = T^{-1}A_1B_1'X_t + \varepsilon_t$$

the local hypothesis can be tested. A_1 and B_1 matrices are (nx(r-r_0)) and (mx(r-r_0)) dimensional respectively, r-r_0>0 and ε_t is the error term. The reduced rank estimators of A and B matrices are found as follows. $\hat{l}_1 \geq \cdots \geq \hat{l}_n$ being the eigenvalues, the $\widehat{\boldsymbol{b}}_{\boldsymbol{1}}, \ldots, \widehat{\boldsymbol{b}}_{\boldsymbol{n}}$ eigenvectors obtained by the solution of the equation below

$$\det(M_{XY}M_{YY}^{-1}M_{YX} - lM_{XX}) = 0$$

$$M_{XX} = T^{-1}\sum_{t=1}^{T} X_t X_t' \quad , M_{XY} = M_{YX}' = T^{-1}\sum_{t=1}^{T} X_t Y_t', \quad M_{YY} = T^{-1}\sum_{t=1}^{T} Y_t Y_t'$$

corresponds to the $\hat{l}_1, \ldots, \hat{l}_n$ eigenvalues. Such that these eigenvalues and eigenvectors provide the equation below:

$$(M_{XY}M_{YY}^{-1}M_{YX} - \hat{l}_j M_{XX})\hat{b}_j = 0$$

In addition the eigenvectors are normalized as follows to reduce the change and to convert the size of the eigenvectors into units:

$$\hat{b}_i' M_{XX} \hat{b}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Later, if the reduced regression estimators of the coefficients $\hat{B} = [\hat{b}_1, \ldots, \hat{b}_r]$ and Y_t are regressed on $\hat{B}'X_t$ using the least square method $\hat{A}$ matrix will be obtained.

Let the reduced rank equity in the Equation 5.2 to be defined as such that the error term a_t be equal to, ε_t and the local alternative hypothesis can be tested. For this purpose Equaiton 5.3 is obtained by multiplying Equation 5.2 by $\hat{A}_\perp'$.

$$\hat{A}_\perp' Y_t = \Phi \hat{U}_t + R\hat{V}_t + N_t \qquad (5.3)$$

$$\hat{U}_t = \hat{B}'X_t, \quad \hat{V}_t = \hat{B}_\perp' X_t, \quad \Phi = \hat{A}_\perp' A \quad \text{ve} \quad N_t = \hat{A}_\perp' \Delta X_t - \hat{A}_\perp' A(\hat{B} - B)' X_t$$

In Equation 5.3 the real value of the coefficient R is zero and the following test statistics is used for testing the alternative hypothesis:

$$LR(r_0) = T\left[logdet\left(\hat{\Sigma}_{NN}^{r}\right) - logdet\left(\hat{\Sigma}_{NN}\right)\right] \tag{5.4}$$

In Equation 5.4, let

$$\hat{\Sigma}_{NN} = T^{-1}\sum_{t=1}^{T}(\hat{A}'_{\perp}Y_t - \hat{\Phi}\hat{U}_t - \hat{R}\hat{V}_t)\,(\hat{A}'_{\perp}Y_t - \hat{\Phi}\hat{U}_t - \hat{R}\hat{V}_t)' \tag{5.5}$$

$$\hat{\Sigma}_{NN}^{r} = T^{-1}\sum_{t=1}^{T}(\hat{A}'_{\perp}Y_t - \hat{\Phi}_r\hat{U}_t)(\hat{A}'_{\perp}Y_t - \hat{\Phi}_r\hat{U}_t)' \tag{5.6}$$

and $\hat{\Sigma}_{NN}$ and $\hat{\Sigma}_{NN}^{r}$ are the estimators of the variance-covariance matrix belonging to the error term in Equation 5.4, respectively R=0 being restricted and $R \neq 0$ being unrestricted. $\hat{\Phi}$ and $\hat{R}$ are the least squares estimators of the parameters in Equation 5.3. $\hat{\Phi}_r$ is the estimator under R=0 restriction.

Theorem 5.1. The $LR(r_0)$ statistics in Equation 5.4 is equal to testing the $H_0: rank(\Upsilon) = r_0$ null hypothesis against $H_1: rank(\Upsilon) = n$ alternative hypothesis for the $Y_t = \Upsilon X_t + \varepsilon_t$ model. (saıkkonen &lutkepohl 1999)

Proof: (saıkkonen & lutkepohl 1999)

If the least squares method is used for the estimation of the parameters in the model in 5.2;

$$Y_t = \hat{\Upsilon}X_t + \hat{\varepsilon}_t \quad t = 1, \dots, T \tag{5.7}$$

will be obtained. The likelihood ratio statistics for the $H_0: rank(\Upsilon) = r_0$ null hypothesis is expressed as below:

$$LR(r_0) = -T\sum_{j=r_0+1}^{n}\log\left(1 + \hat{\lambda}_j\right) \tag{5.8}$$

Here $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ eigenvalueas are obtained by the solution of the equation $\det(\hat{\Upsilon}M_{XX}\hat{\Upsilon}' - \lambda\hat{\Sigma}_{\varepsilon\varepsilon}) = 0$. $\hat{\boldsymbol{v}}_{\boldsymbol{1}}, \dots, \hat{\boldsymbol{v}}_{\boldsymbol{n}}$ normalized eigenvectors correspond to the, $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ eigenvalues. Thus the following equation is obtained:

$$\left(\hat{\Upsilon} M_{XX} \hat{\Upsilon}' - \hat{\lambda}_j \hat{\Sigma}_{\varepsilon\varepsilon}\right) v_j = 0$$

Let the Gaussian maximum likelihood estimator of $B = [b_1, \dots, b_r]$ matrix be $\hat{B} = [\hat{b}_1, \dots, \hat{b}_r]$, the equation below is obtained:

$$\hat{b}_j = \hat{\lambda}_j^{-1/2} \hat{\Upsilon}' \hat{v}_j \qquad j = 1, \dots, n$$

Let $\hat{v} = [\hat{v}_1, \dots, \hat{v}_r]$, $\hat{v}_* = [\hat{v}_{r+1}, \dots, \hat{v}_n]$ and $\hat{B}_* = [\hat{b}_{r+1}, \dots, \hat{b}_n]$, if we multiply Equation 5.7 by $[\hat{v}: \hat{v}_*]'$ under the assumption that r= r_0 null hypothesis is true, we obtain the following equation:

$$\hat{v}' Y_t = \hat{\Lambda}^{1/2} \hat{B}' X_t + \hat{v}' \tilde{\varepsilon}_t$$

$$\hat{v}_*' Y_t = \hat{\Lambda}_*^{1/2} \hat{B}_*' X_t + \hat{v}_*' \tilde{\varepsilon}_t \qquad (5.9)$$

Here let $\hat{\Lambda} = diag[\hat{\lambda}_1, \dots, \hat{\lambda}_r]$, $\hat{\Lambda}_* = [\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_n]$, the likelihood ratio statistics in Equation 5.8 is obtained using $\hat{\Lambda}_*^{1/2}$ in Equation 5.9. This indicates that when the model in Equation 5.7 is multiplied by an appropriate matrix the likelihood ratio statistics can be obtained. If we are to explain this situation in a more comprehensible way, we can write the following equation using Equation 5.9 with the expression $[\hat{\eta}: \hat{\eta}_*]' = [\hat{v}: \hat{v}_*]^{-1}$:

$$Y_t = \hat{\eta} \hat{\Lambda}^{\frac{1}{2}} \hat{B}' X_t + \hat{\eta}_* \hat{\Lambda}_*^{\frac{1}{2}} \hat{B}_*' X_t + \tilde{\varepsilon}_t$$

$$= \hat{A} \hat{B}' X_t + \hat{A}_* \hat{B}_*' X_t + \tilde{\varepsilon}_t \qquad (5.10)$$

Here the expressions $\hat{A} = \hat{\eta} \hat{\Lambda}^{\frac{1}{2}}$ and $\hat{A}_* = \hat{\eta}_* \hat{\Lambda}_*^{\frac{1}{2}}$ are written. In other words, the solution of the model in 5.10 with reduced regression method is equal to the model below:

$$Y_t = AB' X_t + u_t \qquad t = 1, \dots, \mathrm{T}$$

$$u_t = T^{-1} A_1 B_1' X_t + \varepsilon_t$$

Thus we can conclude that, testing the models below

$$Y_t = AB'X_t + a_t \qquad t = 1, \dots, \mathrm{T}$$

$$a_t = T^{-1}A_1B_1'X_t + \varepsilon_t$$

using the following alternative local hypothesis

$$H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

is equal to testing the alternative hypothesis

$$H_1: r = n$$

In the following model:

$$Y_t = \Upsilon X_t + \varepsilon_t$$

In order to obtain the asymptotic properties of the Johansen Trace test statistics under local alternative hypothesis the process should be first order integrated, i.e. $I + \beta'\alpha$ vector's eigenvalues should be limited between the values ∓ 1 (Johansen, 1995). This can be mathematically written follows:

$$|eig(I + \beta'\alpha)| < 1$$

Theorem 5.2 (Johansen 1995)

Let $X^T = \alpha\beta' X_t + \alpha_1\beta_1' X_t + \varepsilon_t$; it has the distribution below:

$$T^{-\frac{1}{2}}\alpha_\perp' X^T_{[Tu]} \rightarrow K(u)$$

Let $X^T_{[Tu]}$ be the stepwise function for $u\epsilon[0,1]$, $K(u)$ is defined as Ornstein-Uhlenbeck process in the literature. Ornstein-Uhlenbeck process is expressed as follows:

$$K(t) = \alpha_\perp' \int_0^t \exp\big(\alpha_1 * \beta_1' * C * (t-u)\big)\, dW(u)$$

W represents the Brownian motion with p dimensional Ω matrix. Also $C = \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha_\perp$. $K(u)$ process provides the stochastic integral equation below:

$$K(u) = \alpha'_\perp W(u) + \alpha'_\perp\alpha_1\beta'_1\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\int_0^u K(s)ds$$

Theorem 5.3 (Johansen 1995)

When the $H_0(r_0): \Pi = \alpha\beta'$null hypothesis is examined under the $H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta'_1$ local alternative hypothesis, if $|eig(I + \alpha'\beta)| < 1$ then the asymptotic distribution of Johansen Trace test is as follows

$$tr\left\{\int_0^1 (dK^*)K^{*\prime}\left[\int_0^1 K^*K^{*\prime}du\right]^{-1}\int_0^1 K^*(dK^*)'\right\} \qquad (5.11)$$

Here K^* is the p-r dimensional Ornstein-Uhlenbeck process, and it provides the equation below:

$$-ab'\int_0^t K^*(u)du + K^*(t) = B(t) \quad t\epsilon[0,1] \qquad (5.12)$$

In Equation 5.12, $B(u)$ is a p-r dimensional Brownian process. Also a and b coefficients are as follows:

$$a = (\alpha'_\perp\Omega\alpha_\perp)^{-1/2}\alpha'_\perp\alpha_1 \quad , \quad b = (\alpha'_\perp\Omega\alpha_\perp)^{-1/2}(\beta'_\perp\alpha_\perp)^{-1}\beta'_\perp\beta_1$$

By calculating the differential of both sides of equation 5.12, the following equation is obtained:

$$d\,K^*(t) = ab'\,K^*(t)dt + dB(t) \qquad (5.13)$$

Especially, equation 5.13 will be used in data generation in the simulation section.

5.2 Johansen Maximum EigenValue Test and its Power Under Local Alternative Hypothesis

The procedure pertaining to the reduced rank regression used for obtaining the asymptotic properties of Johansen Trace test is used to obtain the asymptotic properties of Johansen Maximum EigenValue test in this section. More detailed information can be found in Lüthkepohl, Saikkone & Trenkler (2000). When the reduced regression model

$$Y_t = AB'X_t + a_t \qquad t = 1, \dots, \mathrm{T} \tag{5.14}$$

is considered, here Y_t and a_t are (nx1) dimensional, m>n, X_t is (mx1) dimensional, and A and B matrices are (nxr_0) and (mxr_0) dimensional respectively. If the form of the error term a_t is constituted as below

$$a_t = T^{-1}A_1B_1'X_t + \varepsilon_t$$

The local alternative hypothesis can be tested.

$$H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

Under the local hypothesis the test statistics is

$$LR_{max}(r_0) = -Tlog(1 - \hat{l}_{r_0+1})$$

and $\hat{l}_1 \geq \cdots \geq \hat{l}_n$ eigenvalues are obtained with the solution of the equation below:

$$\det(M_{XY}M_{YY}^{-1}M_{YX} - lT^{-1}M_{XX}) = 0$$

$$M_{XX} = T^{-1}\sum_{t=1}^{T} X_tX_t' \quad , M_{XY} = M_{YX}' = T^{-1}\sum_{t=1}^{T} X_tY_t', \quad M_{YY} = T^{-1}\sum_{t=1}^{T} Y_tY_t'$$

The asymptotic distribution of the $LR_{max}(r_0)$ statistics under the local alternative hypothesis is as follows:

$$LR_{max}(r_0) \xrightarrow{d} l_{max}\left\{\int_0^1 (dK^*)K^{*\prime}\left[\int_0^1 K^*K^{*\prime}du\right]^{-1}\int_0^1 K^*(dK^*)'\right\} \tag{5.15}$$

Here K^* is a p-r dimensional Ornstein-Uhlenbeck process and this process provides the following equation:

$$-ab' \int_0^t K^*(u)du + K^*(t) = B(t) \quad t\epsilon[0,1] \tag{5.16}$$

In Equation 5.15 $B(u)$ is a p-r dimensional Brownian process. Also the a and b coefficients are as follows:

$$a = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \alpha'_\perp \alpha_1 \quad , \quad b = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} (\beta'_\perp \alpha_\perp)^{-1} \beta'_\perp \beta_1$$

By calculating the differential for boths sides of Equation 5.16 the following equation is obtained:

$$d\,K^*(t) = ab'\,K^*(t)dt + dB(t) \tag{5.17}$$

This equation will be used in the simulation section.

5.3 Simulation: Comparison of the Power Functions of Johansen Trace Test and Maximum Eigenvalue Tests Under Local Alternative Hypothesis

Due to the non-standard distributions of the local power function in Equation 5.11, the relative efficiency of the test is not clear. The discrete version of the Ornstein-Uhlenbeck process K(s) is generated by the simulation and the local power is calculated.

Theorem 5.3 (Johansen 1995)

Under $\Pi_T = \alpha\beta' + T^{-1}\alpha_1\beta_1'$ local alternative the asymptotic power only depends on the $f = \beta_1' C \alpha_1 < 0$ and $g = \left(\alpha_1'\alpha_\perp \left(\alpha'_\perp \Omega \alpha_\perp\right)^{-1} \alpha'_\perp \alpha_1\right)(\beta_1' C \Omega C' \beta_1) - (\beta_1' C \alpha_1)^2$ parameters. Here $= \beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha_\perp$.

Using Equation 5.13, the data are generated in the simulation from the process below:

$$\Delta K_t = \frac{1}{T}\alpha_1 \beta_1' K_{t-1} + e_t \qquad e_t \sim i.i.d\ \ N(0, I_n)$$

$$\beta_1' = \begin{cases} 1, & n = 1 \\ (1,0), n = 2 \\ (1,0,0), n = 3 \\ (1,0,0,0), n = 4 \end{cases}$$

$$\alpha_1' = \begin{cases} 1, & n = 1 \\ (f,0), n = 2 \\ (f,g,0), n = 3 \\ (f,g,0,0), n = 4 \end{cases}$$

From these data the following hypotheses are tested using the Johansen Trace test and the Maximum EigenValue tests, and the local powers of the tests obtained for t=30 and 400:

$$H_0(r_0): \Pi = \alpha\beta'$$

$$H_T(r_0): \Pi = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

2000 trial are made for n=1,2,3 and 4. Using the K_t process for both tests, let

$$A_t = T^{-2}\sum_{t=1}^{T} K_{t-1}K'_{t-1} \quad , \quad B_t = T^{-1}\sum_{t=1}^{T} K_{t-1}\Delta K'_t$$

the limit distributions of the trace and maximum eigenvalue test statistics are calculated as below:

$$tr(B'_t A_t^{-1} B_t) \quad , \quad l_{max}(B'_t A_t^{-1} B_t)$$

Using the critical values obtained by G. MacKinon, A.Haug, L.Michelis (1998) for the calculated test statistics in each trial the rejection frequency, i.e. local power is calculated. These calculations are repeated for different time spans and variable numbers, and it is tried to determine which test has better power value for different situations. The results are presented in tables 5.1, 5.2 and figures 5.1 and 5.2.

Table 5. 1 Local Power Values calculated for T=30 for Johansen Trace and Maximum Eigen-Value Tests

n	g		f=0	-1,5	-3	-4,5	-6	-7,5	-9	-10,5	-12	-13,5	-15	-16,5	-18	-19,5	-21
1	-	Tr	0,049	0,068	0,145	0,257	0,361	0,506	0,637	0,765	0,847	0,902	0,936	0,949	0,974	0,989	0,994
		Max															
2	0	Tr	0,029	0,039	0,055	0,076	0,109	0,148	0,216	0,285	0,376	0,447	0,532	0,623	0,697	0,758	0,829
		Max	0,031	0,037	0,051	0,072	0,118	0,144	0,201	0,285	0,384	0,445	0,538	0,630	0,704	0,769	0,842
2	6	Tr	0,601	0,440	0,330	0,304	0,306	0,348	0,378	0,427	0,491	0,568	0,648	0,708	0,765	0,826	0,852
		Max	0,585	0,423	0,319	0,293	0,292	0,331	0,374	0,422	0,502	0,582	0,647	0,709	0,778	0,831	0,866
2	12	Tr	0,930	0,890	0,836	0,789	0,754	0,771	0,760	0,779	0,779	0,806	0,824	0,865	0,884	0,905	0,929
		Max	0,930	0,891	0,840	0,791	0,759	0,772	0,776	0,788	0,784	0,817	0,836	0,869	0,895	0,913	0,935
3	0	Tr	0,033	0,033	0,032	0,034	0,05	0,057	0,095	0,107	0,164	0,199	0,270	0,318	0,393	0,459	0,562
		Max	0,027	0,029	0,034	0,036	0,052	0,062	0,101	0,135	0,180	0,229	0,309	0,377	0,461	0,520	0,618
3	6	Tr	0,474	0,286	0,185	0,156	0,14	0,143	0,173	0,205	0,232	0,298	0,348	0,405	0,465	0,544	0,593
		Max	0,471	0,285	0,185	0,159	0,149	0,158	0,179	0,225	0,278	0,320	0,382	0,469	0,515	0,599	0,657
3	12	Tr	0,838	0,761	0,682	0,586	0,535	0,497	0,470	0,468	0,519	0,524	0,563	0,612	0,659	0,683	0,751
		Max	0,846	0,778	0,706	0,615	0,553	0,535	0,521	0,529	0,571	0,595	0,619	0,687	0,711	0,752	0,795
4	0	Tr	0,015	0,015	0,022	0,016	0,022	0,026	0,037	0,049	0,067	0,091	0,136	0,170	0,198	0,276	0,311
		Max	0,023	0,019	0,028	0,027	0,034	0,034	0,047	0,077	0,091	0,129	0,191	0,236	0,279	0,370	0,442
4	6	Tr	0,372	0,172	0,113	0,078	0,067	0,081	0,079	0,104	0,109	0,134	0,154	0,208	0,249	0,299	0,364
		Max	0,379	0,189	0,109	0,089	0,095	0,100	0,102	0,124	0,157	0,207	0,217	0,300	0,352	0,407	0,485
4	12	Tr	0,759	0,659	0,528	0,396	0,333	0,317	0,288	0,283	0,302	0,305	0,332	0,369	0,428	0,443	0,513
		Max	0,795	0,700	0,578	0,468	0,406	0,383	0,359	0,368	0,399	0,402	0,443	0,493	0,542	0,586	0,640

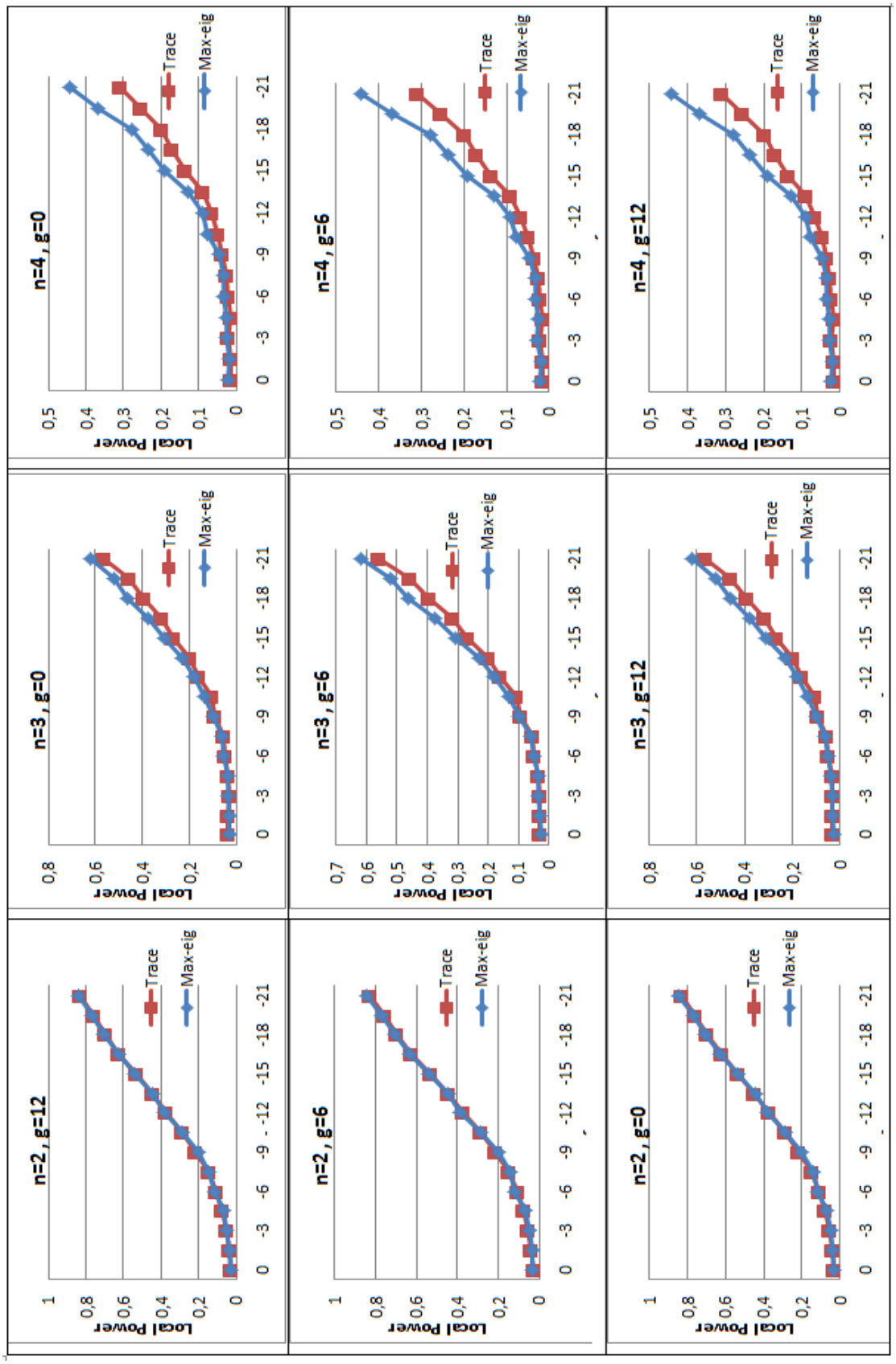

Figure 5.1 Local Power Values variably calculated for T=30 for Johansen Trace and Maximum Eigen-Value Tests

Table 5.2 Local Power Values calculated for T=400 for Johansen Trace and Maximum Eigen-Value Tests

n	g		f=0	-1,5	-3	-4,5	-6	-7,5	-9	-10,5	-12	-13,5	-15	-16,5	-18	-19,5	-21
1	-	Tr	0,045	0,085	0,134	0,223	0,351	0,473	0,604	0,734	0,822	0,900	0,942	0,968	0,986	0,994	0,997
		Max															
2	0	Tr	0,054	0,051	0,062	0,075	0,102	0,128	0,161	0,198	0,268	0,325	0,389	0,496	0,570	0,632	0,684
		Max	0,056	0,046	0,058	0,063	0,092	0,122	0,145	0,2	0,259	0,321	0,394	0,493	0,560	0,629	0,692
2	6	Tr	0,598	0,423	0,327	0,278	0,269	0,284	0,308	0,352	0,387	0,469	0,531	0,569	0,629	0,699	0,751
		Max	0,589	0,421	0,314	0,269	0,252	0,279	0,284	0,337	0,394	0,459	0,524	0,571	0,633	0,700	0,765
2	12	Tr	0,941	0,889	0,849	0,791	0,757	0,713	0,715	0,728	0,728	0,731	0,779	0,799	0,824	0,858	0,886
		Max	0,931	0,895	0,841	0,788	0,739	0,711	0,717	0,727	0,727	0,734	0,776	0,820	0,833	0,849	0,895
3	0	Tr	0,047	0,048	0,047	0,056	0,057	0,081	0,094	0,113	0,124	0,153	0,192	0,218	0,259	0,293	0,349
		Max	0,042	0,043	0,048	0,052	0,061	0,068	0,087	0,085	0,124	0,138	0,174	0,221	0,260	0,298	0,348
3	6	Tr	0,517	0,295	0,209	0,153	0,154	0,154	0,16	0,171	0,198	0,218	0,256	0,281	0,314	0,362	0,392
		Max	0,475	0,269	0,185	0,145	0,145	0,141	0,142	0,164	0,172	0,211	0,259	0,293	0,320	0,363	0,423
3	12	Tr	0,881	0,781	0,684	0,593	0,523	0,462	0,426	0,429	0,433	0,432	0,436	0,458	0,490	0,518	0,551
		Max	0,865	0,778	0,681	0,610	0,513	0,467	0,447	0,395	0,437	0,446	0,486	0,473	0,505	0,549	0,589
4	0	Tr	0,047	0,038	0,045	0,038	0,050	0,061	0,058	0,077	0,079	0,103	0,121	0,132	0,129	0,175	0,193
		Max	0,052	0,047	0,045	0,044	0,053	0,057	0,061	0,080	0,076	0,095	0,112	0,142	0,141	0,184	0,198
4	6	Tr	0,425	0,252	0,154	0,119	0,098	0,096	0,106	0,104	0,112	0,129	0,141	0,164	0,179	0,206	0,219
		Max	0,411	0,229	0,127	0,105	0,092	0,095	0,097	0,090	0,115	0,116	0,137	0,176	0,173	0,202	0,233
4	12	Tr	0,825	0,711	0,575	0,449	0,378	0,341	0,302	0,285	0,276	0,279	0,276	0,285	0,286	0,318	0,346
		Max	0,829	0,703	0,579	0,452	0,376	0,347	0,302	0,278	0,270	0,292	0,297	0,296	0,333	0,339	0,390

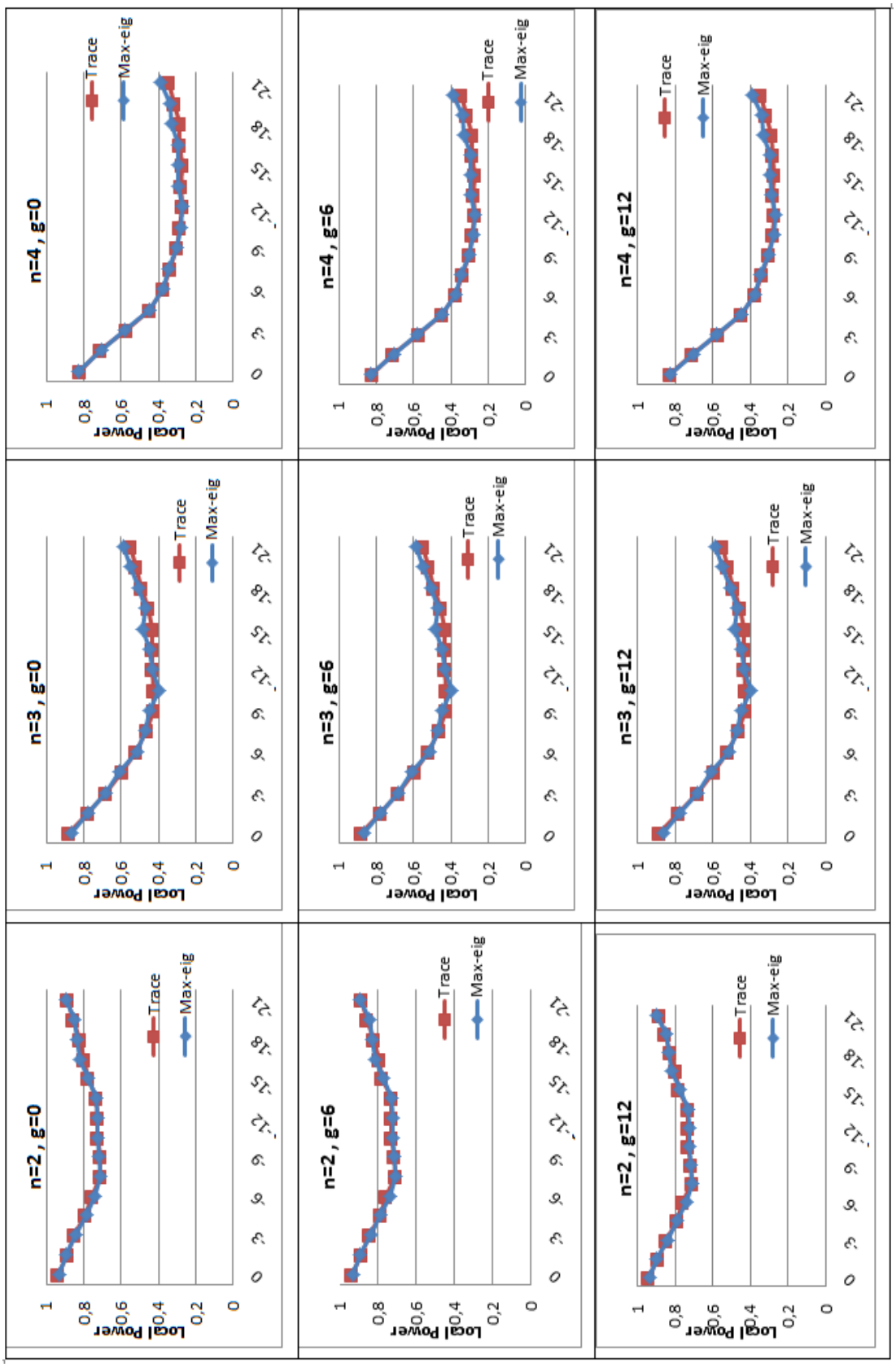

Figure 5.2 Local Power Values calculated for T=400 for Johansen Trace and Maximum Eigen-Value Tests

With regard to the results obtained in the simulations, the power value decreases for both tests as the number of variables increases. Also, for smaller time span (T=30), the maximum eigenvalue test gives better results than trace test, as the number of the variables increases. On the other hand, for greater time span, both tests have the same power values.

CHAPTER SIX
CONCLUSION

The results obtained from this study are listed below:

- The asymptotic distributions of the $\breve{\mathbf{\Pi}}$, $\breve{\boldsymbol{\alpha}}$ and $\breve{\boldsymbol{\beta}}$ maximum likelihood estimators for the $\Delta\mathbf{X_t} = \mathbf{\Pi X_{t-1}} + \mathbf{a_t} = \boldsymbol{\alpha\beta}'\mathbf{X_{t-1}} + \mathbf{a_t}$ $t = 1,2,\ldots$ model is the same as the asymptotic distributions of and they present properties pertaining to normal distribution. In addition to this, knowing the cointegration matrix $\boldsymbol{\beta}$ does not affect the asymptotic properties of the estimators of $\mathbf{\Pi}$ matrix.
- $A = \begin{bmatrix} \rho & \theta \\ 0 & \alpha \end{bmatrix}$ being the coefficients matrix, unit root for the cointegrated $\mathbf{X_t} = \mathbf{AX_{t-1}} + \mathbf{a_t}$ model is obtained in two different cases. These cases are obtained by selecting either $\rho = 1$ and $\alpha < 1$ or $\alpha = 1$ and $\rho < 1$. For $\rho = 1$ and $\alpha < 1$ case, the mean square error value of the (α, ρ) parameters yields better results.
- The local powers for the $\alpha = 0.05$ confidence level were found for the Johansen Trace and Maximum Eigenvalue tests, used for determining the presence of cointegration. For the comparison of these local powers the increasing time intervals and the number of variables in the model are critically selected. In case the time intervals are wide, the local power values for both test methods are equal. When narrower time intervals are selected, presence of one or two variables in the system yields equal local power values; however, when there are three or more variables in the system maximum eigenvalue test gives better results than trace test.
- In this study, there is not any constant, variable and/or linear trend in the system. All analyses conducted can be repeated considering this situation. Besides, only the long term relations of the economic variables are handled in the system (simple model). In addition to the long term relations, a model should be constructed including short term relations and the results should be re-evaluated.

REFERENCES

Akdi, Y. (2010). *Zaman serileri analizi (Birim kökler ve kointegrasyon).*Ankara: Gazi Kitabevi.

Engle, R. F. & Granger, C.W.J. (1987). *Cointegration for the applied economist*, London: The Mac Millan Press Ltd.

Hamilton, J.D. (1994). *Time series analysis*, New Jersey: Princeton University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control .12* , 231-254.

Johansen,S. (1995). *Likelihood-based inference in cointegrated vector-autoregressive models.* Oxford: Oxford University Press.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models.*Econometrica. 59*, 1551–80

Johansen, S. & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics. 52*, 169–210.

Lütkepohl, H. (2006*). New introduction to multiple time series analysis*. Berlin: Springer.

Lütkepohl, H. G. & Saikkonen, P. A., & Trenkler, C. E. (2000). Maximum Eigenvalue versus Trace tests for the cointegration rank of a VAR process. *Econometrics Journal 4*, 287–310.

McManus, A. (1991). Who invented local power analysis? .*Econometric Theory*. *7*, 265-268.

Saikkonen, P. & H. Lutkepohl (1999). Local power of likelihood ratio tests for the cointegrating rank of aVAR process.*Econometric Theory*. *15*,50–78.

Wei, W., (2006). *Time series analysis: univariate and multivariate methods*. Pearson:United States.

APPENDIX 1

MATLAB CODES

```
clear
   max=0;
   tra=0;
for i=1:2000
   i

Spec=vgxset('AR',{[0.9475,0,0,0;0.03,1,0,0;0,0,1,0;0,0,0,1]},'Q',[1,0,0,0;0,1,0,0;0,0,
1,0;0,0,0,1]);
   Y = vgxsim(Spec,400);
   farky=diff(Y);
   lagy1=Y(1:399,1);
   lagy2=Y(1:399,2);
   lagy3=Y(1:399,3);
   lagy4=Y(1:399,4);
   lagy=[lagy1,lagy2,lagy3,lagy4];
   sonucat=(lagy'*lagy)/400^2;
   sonucbt=(lagy'*farky)/400;
   tr=trace(sonucbt'*(sonucat^-1)*sonucbt);
   [V,D] = eig(sonucbt'*(sonucat^-1)*sonucbt);
   x=sort([D(1,1),D(2,2),D(3,3),D(4,4)]);
   sontr(i,1)=tr;
   sonmax(i,1)=x(1,4);

   if  x(1,4)>24,16
     max=max+1;
   end

   if   tr>40.17
      tra=tra+1;
   end
end
   powmax=max/2000
   powtra=tra/2000
```