ASPECTS OF BIVARIATE TIME SERIES

by

SOLLY MATSHONISA SEELETSE

submitted in part fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROFESSOR R MARKHAM

November 1994
I declare that *Aspects of Bivariate Time Series* is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

SM SEELETSE

20/11/14
ACKNOWLEDGEMENTS

The following people have been most helpful in my studies and they have made it possible for me to reach this stage of my studies:

- Prof R Markham for his guidance from the time I was busy for the Hons degree, and for the choice of other modules/projects for this degree. Most of all, his guidance for this dissertation was excellent. It's hard to explain it.

- Mrs Anet Coetzee who agreed to do the typing even at the late stage she was requested.

- My mother, mme Malebo and brother, Funky for their support especially when I was down.
NOTATION

A: m×n : matrix of order m×n
A(m×n) :

a : vector
A' : transpose of matrix A
|A| : determinant of

det(A) :
rank(A) : rank of matrix A
diag(λ1, ..., λn) : diagonal matrix of diagonal elements λ1, ..., λn

D, : p×p identity matrix
Ip : p×p zero matrix (zero matrix)
O : zero vector
A > B : A - B is positive definite
A ≥ B : A - B is positive semidefinite
A\B : set A with elements of set B removed*

: approximately equal to
: distributed as
ML : maximum likelihood
exp(x) : e^x
\chi^2_v : chi-square with v degrees of freedom
N(µ, Σ) : normal distribution with mean vector µ and covariance matrix Σ

mgf : moment generating function

* Sets will be represented by events to avoid common notation with matrices.
Normal density function (multivariate)

\[ f(z) = \frac{1}{\sqrt{2\pi}^{p/2}} \frac{1}{\left|\Sigma\right|^{1/2}} \exp \left[ -\frac{(z - \mu)^\top \Sigma^{-1} (z - \mu)}{2} \right] \]

mgf of \( X_v^2 \)

\[ (1 - 2t)^{-v/2} \]

(mgf is unique for every distribution if it exists)
# CONTENTS

## SUMMARY

<table>
<thead>
<tr>
<th>Summary</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vii</td>
</tr>
</tbody>
</table>

## CHAPTER 1: INTRODUCTION

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Problems with exponential smoothing</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Classes of statistical models</td>
<td>3</td>
</tr>
<tr>
<td>1.3 What lies ahead?</td>
<td>3</td>
</tr>
</tbody>
</table>

## CHAPTER 2: BIVARIATE EXPONENTIAL SMOOTHING ALGORITHMS, THE SUGGESTED STATISTICAL MODELS AND THE STATE SPACE MODEL

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Bivariate exponential smoothing and suggested bivariate ARMA processes</td>
<td>5</td>
</tr>
<tr>
<td>2.1.1 Simple bivariate exponential smoothing and bivariate IMA(1,1)</td>
<td>5</td>
</tr>
<tr>
<td>2.1.2 Bivariate generalisation of Holt's algorithm and the local level constant slope model</td>
<td>8</td>
</tr>
<tr>
<td>2.1.3 Bivariate damped trend algorithm and ARIMA(1,1,2)</td>
<td>12</td>
</tr>
<tr>
<td>2.2 Bivariate exponential smoothing and suggested bivariate structural processes</td>
<td>15</td>
</tr>
<tr>
<td>2.2.1 Simple bivariate exponential smoothing algorithm suggests bivariate random walk plus noise</td>
<td>17</td>
</tr>
<tr>
<td>2.2.2 Bivariate generalization of Holt's exponential smoothing algorithm suggests bivariate local linear trend</td>
<td>18</td>
</tr>
<tr>
<td>2.2.3 Bivariate structural model suggested by the bivariate damped trend algorithm</td>
<td>19</td>
</tr>
<tr>
<td>2.2.2 Remarks</td>
<td>21</td>
</tr>
<tr>
<td>2.3 State space model</td>
<td>26</td>
</tr>
<tr>
<td>2.3.1 State space representation (SSR)</td>
<td>27</td>
</tr>
<tr>
<td>2.3.2 Kalman filter</td>
<td>28</td>
</tr>
<tr>
<td>2.3.3 Kalman forecasts</td>
<td>31</td>
</tr>
<tr>
<td>2.3.4 Formulae for point forecasts</td>
<td>34</td>
</tr>
<tr>
<td>2.3.5 Formulae for forecast MSE</td>
<td>35</td>
</tr>
</tbody>
</table>

## CHAPTER 3: CAUSALITY (AND FEEDBACK) OF ARIMA MODELS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Definitions and assumptions</td>
<td>38</td>
</tr>
<tr>
<td>3.2 More definitions: forecasting concepts</td>
<td>50</td>
</tr>
<tr>
<td>3.3 Grangers concept of causality</td>
<td>52</td>
</tr>
<tr>
<td>3.3.1 Characterization of Granger-causality in bivariate time series</td>
<td>54</td>
</tr>
<tr>
<td>3.3.2 Characterization of instantaneous causality</td>
<td>61</td>
</tr>
</tbody>
</table>
3.4 Granger-causality of stable ARMA(p,q) 67
3.4.1 ARMA(1,1) process as infinite order MA process 68
3.4.2 Bivariate ARMA(p,q) as prediction error MA representation 73
3.4.3 Granger-causality of bivariate ARMA(p,q) 76

CHAPTER 4: COINTEGRATION 80
4.1 Cointegrated processes 80
4.2 Cointegrated bivariate ARMA processes 87
4.3 Granger-causal cointegrated bivariate ARMA processes 93

CHAPTER 5: FORECASTING: BIVARIATE INTEGRATED AND COINTEGRATED ARMA PROCESSES 98
5.1 Bivariate IMA(1,1) process 99
5.1.1 Point forecasts 100
5.1.2 Forecast (or minimum) mean square error matrices (MMSE) 101
5.1.3 Granger-causal IMA(1,1) 104
5.2 Bivariate IMA(2,2) 105
5.2.1 Point forecasts 106
5.2.2 MMSE matrices 111
5.2.3 Granger-causal IMA(2,2) 114
5.3 Bivariate SUTSE ARIMA(1,1,2) model 115
5.4 Forecast regions 120
5.5 Discussion 131

CHAPTER 6: COINTEGRATED BIVARIATE STRUCTURAL TIME SERIES MODELS 133
6.1 Dynamic factor analysis 133
6.1.1 Bivariate random walk plus noise with drift 133
6.1.2 Derivation of common level common slope model 135
6.1.3 Common slopes 137
6.2 Forecasts 138
6.2.1 Forecasts for common level, common slope model 139
6.2.2 Forecasts for the common slope model 144
6.3 Discussion 150
6.3.2 Point forecasts and forecast regions 151

CHAPTER 7: BIVARIATE AUTOREGRESSIONS WITH STOCHASTIC LEVELS/TRENDS 152
7.1 Additive model 158
7.2 Bivariate autoregressive model 164
7.3 Dynamic model 169
7.4 Closing remarks 173
# Chapter 8: Common Levels/Trends and Cointegration of Autoregressions

- **8.1 Additive model**
- **8.2 Bivariate autoregressive trends**
- **8.3 Dynamic model**
- **8.4 Remark**

# Chapter 9: Forecasting Bivariate Structural Time Series Models

- **9.1 Point forecasts and forecast MSE’s**
  - **9.1.1 Random walk plus noise**
  - **9.1.2 Bivariate local linear trend**
  - **9.1.3 Bivariate damped trend model**
  - **9.1.4 Discussion**
- **9.2 Forecast regions**
  - **9.2.1 Random walk plus noise**
  - **9.2.2 Bivariate local linear trend**
  - **9.2.3 Damped trend model**

# Chapter 10: Comparison of (Bivariate) ARIMA and Structural Models

- **10.1 Generalisations of exponential smoothing by classes of statistical models**
- **10.2 Cointegration properties**
- **10.3 Bivariate autoregressions**
- **10.4 Conclusion**

# Appendix A

# Appendix B

# Bibliography
Exponential smoothing algorithms are very attractive for the practical world such as in industry. When considering bivariate exponential smoothing methods, in addition to the properties of univariate methods, additional properties give insight to relationships between the two components of a process, and also to the overall structure of the model.

It is important to study these properties, but even with the merits the bivariate exponential smoothing algorithms have, exponential smoothing algorithms are nonstatistical/nonstochastic and to study the properties within exponential smoothing may be worthless.

As an alternative approach, the (bivariate) ARIMA and the structural models which are classes of statistical models, are shown to generalize the exponential smoothing algorithms. We study these properties within these classes as they will have implications on exponential smoothing algorithms.

Forecast properties are studied using the state space model and the Kalman filter. Comparison of ARIMA and structural model completes the study.

**KEYWORDS**
Exponential smoothing algorithms, bivariate ARMA models, bivariate structural models, State space models, Kalman filter, Granger-causality, cointegration, point forecasts, forecast regions, autoregressions, common factors.
CHAPTER 1

INTRODUCTION

Jones (1966: 241)
Exponential smoothing algorithms have found numerous industrial applications for a number of reasons, some of which will be given shortly. This smoothing algorithm is described as: "In its updating methods, update is a weighted average of past values of an observed process which gives decreasing weights to past values; and recent observations receive more weight". Computation using this method is also easy.

Jones (1966: 242) states that the mathematical model for which exponential smoothing gives optimum prediction is a special case of some linear model, the statistical importance of this theory of exponential smoothing is that linear regression is a special case of the theory, and therefore it is useful for updating estimates of regression coefficients as more data become available.

Newbold and Bos (1990: 158-9)
Exponential smoothing is a family of algorithms, not a unique forecasting procedure called exponential smoothing. A common approach in all exponential techniques is "least weights for oldest observations"; and each of the several alternative approaches is based on the assumptions about characteristics of time series of interest. The algorithms can only be justified for use now on their previous successes in applications. In fact they are chosen intuitively. Authors stress that "they are ad hoc approaches to problems, and no formal model is fitted on them to explain the behaviour of an observed time series". Many authors have nonetheless "shown that particular exponential algorithms do yield optimal forecasts for time series generated by specific models". The choice of a specific smoothing algorithm is without formal statistical guidelines or as authors say, "the choice ... is arbitrary". Authors claim that these procedures were "developed for routine sales forecasting", and that these algorithms have shown to be "inexpensive, easy to apply and have a successful track record".
1.1 PROBLEMS WITH EXPONENTIAL SMOOTHING

Harvey (1989: 24-5), Newbold and Bos (1990: 159)

We have said that the choice of a suitable exponential technique is arbitrary. There is therefore no formal procedure to evaluate our approach. In this case human judgement is used and not statistical judgement, then the exponential smoothing procedures are open to human bias.

Harvey (1989: 25), Newbold and Bos (1990: 160)

Another problem is the weights given to the observations. There is a problem in choosing suitable weights such that \( \alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 + ... = 1 \) because there are many such \( \alpha \)'s. It becomes an even worse problem when extended to bivariate case of

\[
A = \begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix}
\]

which are required to obey the rules

1. \( A + A(I-A) + A(I-A)^2 + ... = I \)

and

2. \( A(I-A)^i > A(I-A)^j \) for \( i < j \)

There is no such unique \( A \), in fact we extended the unsolved problem of \( \alpha \) to four \( \alpha \)'s, which is worse.

Lastly, the choice of a suitable method. From previous sections we read off that "they are ad hoc approaches to problems, no formal model is fitted on them to explain the behaviour of an observed time series". It is explained by the phrase, "algorithm choice is without formal statistical guidelines" as implied by Newbold and Bos (1990: 159), and this means that the methods are themselves nonstochastic/nonstatistical.

Of course even with these problems, exponential techniques have their own merits as we have explained. We take off by associating these algorithms with classes of statistical models.
1.2 CLASSES OF STATISTICAL MODELS

Harvey (1984: 245)

The problem of choice on method and weight, A, simply means trial and error because if results are not intuitively appealing for the analyst, he may change them. This leads to trial-and-error problem. Our approach is therefore to seek classes of statistical models, then express (or approximate) exponential algorithms as particular forecasting techniques in such classes.

Harvey (1989: 12)

ARIMA models are so popular that statisticians consider it general that every statistical model can be expressed in ARIMA form. In univariate case, Harvey (1984: 257) has outlined the relationship between ARIMA models and exponential techniques.

Harvey (1989: 31)

Recently, structural models have been impressive. Amongst other reasons, in univariate methods, random walk plus noise and local linear trend from structural models are associable with simple exponential and Holt's linear trend of exponential.

Harvey (1984: 257, 279; 1989: 18-9, 75, 656-6)

In the "competition" of (univariate) structural models and ARIMA models, structural models seemed to have emerged a slight winner in that all results analysed through structural models were no worse than any from ARIMA when the reverse is not true. ARIMA seemed to have done well in model building while structural models excelled in forecasting.

We now investigate (for bivariate case) which of structural models and ARIMA can best represent exponential smoothing family of algorithms.

1.3 WHAT LIES AHEAD?

All models are in bivariate form except when stated otherwise.

We derive the relationships between bivariate exponential smoothing and each of ARIMA and structural models in Chapter 2. The chapter is completed by introducing the state space model and derivations of Kalman forecasts.
Causal relationships of ARIMA models are discussed in Chapter 3. We derive also the conditions of these relationships.

Cointegration is discussed in Chapter 4. Unlike causal relationships which is possible even on stationary processes, cointegration is defined only on certain nonstationary processes. Movements which are common in the two components are of interest.

In Chapter 5 forecasts and forecast regions of cointegrated ARMA processes are discussed. Graphical illustrations are given.

Structural models are discussed in Chapters 6 and 9. We introduce cointegrated structural models in Chapter 6 with forecast properties given and illustrated graphically. Chapter 9 treats forecasting in general, of structural models.

Chapters 7 and 8 discuss autoregressions. Three ways of including autoregressions in a model are discussed. In Chapter 7 we illustrate using a numerical example and graphical illustrations. In Chapter 8 we discuss cointegration of autoregressions.

Chapter 10 concludes the study by comparing ARIMA and structural models on the basis of what has been discussed in other chapters.
CHAPTER 2

BIVARIATE EXPONENTIAL SMOOTHING
ALGORITHMS, THE SUGGESTED STATISTICAL
MODELS AND THE STATE SPACE MODEL

Harvey (1989: 23-5)
Exponential smoothing algorithms are useful in practice, but unfortunately
they are nonstochastic (nonstatistical in nature), hence no statistical inference
can be performed on them. They are nevertheless, ad hoc "mechanical"
algorithms which can be used for the calculation of point forecasts. We are
led to the question: Are there classes of statistical models suggested by these
algorithms? A response to this question will be by expressing these
algorithms in classes of statistical models, namely: bivariate ARIMA and
bivariate structural models.

2.1 BIVARIATE EXPONENTIAL SMOOTHING AND THE SUGGESTED
BIVARIATE ARIMA PROCESSES
Bivariate exponential smoothing algorithms are appealing generalizations of
univariate exponential smoothing algorithms. As before, the bivariate
exponential smoothing algorithms are deterministic, that is, no stochastic
components at all. We recall that the simple exponential algorithm has
nonstochastic recurrence form

\[ m_t = az_t + (1-a)m_{t-1} \]

which is a weighted average of two estimates of the current level, namely the
present observation \( x_t \) and the previous estimate \( m_{t-1} \). The smoothing
constant \( a \), satisfies \( 0 < a < 1 \).
A bivariate generalization is

\[
\begin{bmatrix}
m_{1t} \\
m_{2t}
\end{bmatrix}
= \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
+ \begin{bmatrix}
1 - a_{11} & -a_{12} \\
-a_{21} & 1 - a_{22}
\end{bmatrix}
\begin{bmatrix}
m_{1,t-1} \\
m_{2,t-1}
\end{bmatrix}
\] (2.1.1)

where the smoothing parameter

\[
A = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\]

is now a matrix.

In univariate exponential smoothing, the older an estimate the less it contributes in the current estimation because \((1-a)^k\) (which is a "contributor") becomes small for large \(k\). This can be seen by writing \(m_t\) as

\[
m_t = ax_t + a(1-a)x_{t-1} + a(1-a)^2x_{t-2} + \ldots + a(1-a)^kx_{t-k} + \ldots
\]

A bivariate analogue of \((1-a)^k\) decreasing for increasing \(k\) is \((I_2-A)^k\) approaching 0 as \(k\) increases. Lütkepohl (1991: 10/11) explains this case, and concludes that this occurs only when all eigenvalues of \(A\) are less than one in absolute value. This is to say that (cf Lütkepohl (1991: 456) Rule 7)

\[
|I - Az| \neq 0 \quad \text{for} \quad |z| \leq 1
\] (2.1.2)

which is called the stability condition.

As in the univariate case coefficients on the right-hand side of (2.1.1) satisfy

\[A + (I - A) = I.\]

To show relationships suggested in the heading of this section, we start from the error-correction form. The following approach is given in Newbold & Bos.
(1990: 165) for univariate case and was extended to bivariate with the help of Prof. R. Markham.

Starting from simple bivariate exponential smoothing algorithm (2.1.1), suppose information is available up to time $t-1$, then the forecast for the next observation is denoted and given by

$$\begin{bmatrix} x_{1,t|t-1} \\ x_{2,t|t-1} \end{bmatrix} = \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix}$$

Once $\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$ is known, the one-step ahead forecast error vector is

$$\begin{bmatrix} e_{1,t|t-1} \\ e_{2,t|t-1} \end{bmatrix} = \begin{bmatrix} x_{1t} - m_{1,t-1} \\ x_{2t} - m_{2,t-1} \end{bmatrix}$$

(2.1.3)

With (2.1.3) in mind, we rearrange (2.1.1) to get

$$\begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} = \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_{1,t|t-1} \\ e_{2,t|t-1} \end{bmatrix}$$

which, using (2.1.3) we obtain the error-correction form

$$\begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} = \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_{1,1|t-1} \\ e_{2,1|t-1} \end{bmatrix}$$

(2.1.4)

We define the backshift operator $L$.

**Definition (Backshift operator)**

The operator $L$ is called the backshift or lag operator with the property

$$L^k x_t = x_{t-k}, \quad k = 1, 2, \ldots$$

$$L^0 = 1$$
The operator is irrespective of whether univariate or bivariate is in question, in bivariate the same operator gives

\[ L^k [x_{1t}] = [x_{1,t-k}], \quad k = 1, 2, \ldots \]

\[ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \]

\[ L^0 = I_2. \]

2.1.1 Simple bivariate exponential smoothing and bivariate IMA(1,1)

We show that simple bivariate exponential smoothing suggests bivariate IMA(1,1).

Harvey (1989: 25-6), Newbold & Bos (1990: 165-6, 345-55)

The simple bivariate exponential smoothing algorithm assumes that the bivariate vector \( x_{1t} \) is a sum of two components:

\[ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \]

- a level \( \begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} \) which is a function of time
- an error component.

The one-step ahead forecast error vector \( \begin{bmatrix} e_{1,t|t-1} \\ e_{2,t|t-1} \end{bmatrix} \) is an estimate of the error component which occurs at time \( t \).

We use the above suggestion about \( x_t \) and by rewriting (2.1.3) with \( x_t \) as subject of formula we obtain

\[ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t|t-1} \\ e_{2,t|t-1} \end{bmatrix} \]
We apply the operator $I - L$ to both sides of this equation to obtain

$$
\begin{bmatrix}
(I-L)[x_{1t}] \\
[I-L][x_{2t}]
\end{bmatrix} =
\begin{bmatrix}
(I-L)[m_{1,t-1}] \\
(I-L)[m_{2,t-1}]
\end{bmatrix} +
\begin{bmatrix}
(I-L)[e_{1,t|t-1}] \\
[I-L][e_{2,t|t-1}]
\end{bmatrix}
\tag{2.1.5}
$$

We write (2.1.4) using operator $L$, it becomes

$$
\begin{bmatrix}
(I-L)[m_{1t}] \\
[I-L][m_{2t}]
\end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
e_{1,t|t-1} \\
e_{2,t|t-1}
\end{bmatrix}
$$

By substituting $t-1$ for $t$ in this equation and noting the identity

$$(I-L)L = L(I-L)$$

then

$$
\begin{bmatrix}
(I-L)[m_{1,t-1}] \\
[I-L][m_{2,t-1}]
\end{bmatrix} = (I-L)[m_{1t}] \\
\begin{bmatrix}
[I-L][m_{2t}]
\end{bmatrix}
$$

which we substitute into (2.1.5) to obtain

$$
\begin{bmatrix}
(I-L)[x_{1t}] \\
[I-L][x_{2t}]
\end{bmatrix} =
\begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
e_{1,t|t-1} \\
e_{2,t|t-1}
\end{bmatrix} +
\begin{bmatrix}
(I-L)[e_{1,t|t-1}] \\
[I-L][e_{2,t|t-1}]
\end{bmatrix}
$$

which we write as
We still have not assumed any randomness in the above components. As a basis for discussion of random processes we need to understand a (bivariate) white noise process, which we define below.

**Definition (White Noise)**

A zero mean white noise bivariate process is a bivariate random vector

\[
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

with the following properties

1. **mean vector** \( \mathbf{0} \),

\[
\mathbf{E}\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

2. **constant covariance matrix** \( \Sigma_{\varepsilon\varepsilon} \) given by

\[
\Sigma_{\varepsilon\varepsilon} = \begin{bmatrix}
\sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\
\sigma_{21,\varepsilon} & \sigma_{22,\varepsilon}
\end{bmatrix}
\]

where

\[
\sigma_{ij,\varepsilon} = \begin{cases}
\text{cov}(\varepsilon_{1t}, \varepsilon_{jt}), & i \neq j, \ i, j = 1, 2 \\
\text{var}(\varepsilon_{it}), & i = j, \ i = 1, 2
\end{cases}
\]

3. processes \( \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \) and \( \begin{bmatrix} \varepsilon_{1\tau} \\ \varepsilon_{2\tau} \end{bmatrix} \) at different time periods \( t \neq \tau \), are uncorrelated. That is

\[
\text{Cov}\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{1\tau} \\
\varepsilon_{2t} \\
\varepsilon_{2\tau}
\end{bmatrix} = \begin{bmatrix}
\mathbf{E}(\varepsilon_{1t}\varepsilon_{1\tau}) & \mathbf{E}(\varepsilon_{1t}\varepsilon_{2\tau}) \\
\mathbf{E}(\varepsilon_{1\tau}\varepsilon_{1t}) & \mathbf{E}(\varepsilon_{1\tau}\varepsilon_{2\tau}) \\
\mathbf{E}(\varepsilon_{2t}\varepsilon_{1t}) & \mathbf{E}(\varepsilon_{2t}\varepsilon_{2\tau})
\end{bmatrix}
\]
It will simply be called white noise (process). We note an important property that the covariance at all time periods \( t \) is constant (or time invariant).

The association between simple bivariate exponential smoothing is established by assuming that \( [e_{1,t-1} \ e_{2,t-1}] \) of (2.1.6) is a random vector as follows:

Suppose that \( [e_{1,t-1} \ e_{2,t-1}] \) is a bivariate white noise process which we denote by \( [\varepsilon_{1t} \ \varepsilon_{2t}] \), a random vector.

Define

\[
\begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{21} & \theta_{22}
\end{bmatrix} = \begin{bmatrix}
1-a_{11} & -a_{12} \\
-a_{21} & 1-a_{22}
\end{bmatrix}
\]

By substituting these in (2.1.6) we are led to a bivariate IMA(1,1), specifically (2.1.6) suggests the bivariate IMA(1,1) process

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} - \begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix} = \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} - \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{21} & \theta_{22}
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1,t-1} \\
\varepsilon_{2,t-1}
\end{bmatrix}
\]

(2.1.7)

Remarks

(1) The recurrence form (2.1.1) seems to imply that \( x_{2t} \) causes/influences \( m_{1t} \) (as \( a_{12} \neq 0 \)) and \( x_{1t} \) causes \( m_{2t} \) (as \( a_{21} \neq 0 \)). The error-correction form (2.1.6) and bivariate IMA(1,1) process (2.1.7) show that this is not the case. The only dependence between \( x_{1t} \) and \( x_{2t} \) is through the errors (through coefficients \( a_{ij} \) and \( \theta_{ij} \)). We define the previous statement.
Definition (SUTSE)

A bivariate SUTSE process is a bivariate time series \((x_{1t} \ x_{2t})'\) in which the components \(x_{1t}\) and \(x_{2t}\) are related only through their error components.

SUTSE is short for seemingly unrelated time series equations.

The bivariate IMA(1,1) is an example of a SUTSE.

(2) Existence of \((I - \Theta)^{-1}\) which is called invertibility, and estimation of the \(\theta_{ij}\) (for example ML estimators are possible especially when assuming bivariate normality), imply estimation of \(\Theta\). If this is given, then \(A\) of simple bivariate exponential smoothing is given by \(I - \Theta\), where

\[
\Theta = \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{21} & \theta_{22}
\end{bmatrix}
\]

We will discuss the question of \(x_{1t}\) causes \(x_{j_0}\) in Chapter 3.

2.1.2 Bivariate generalization of Holt's algorithm and local constant level, local constant slope model


The recurrence form of bivariate Holt's algorithm is

\[
\begin{align*}
[m_{1t}] &= [a_{11,1} \ a_{12,1}] [z_{1t}] + [1-a_{11,1} \ -a_{12,1}] [m_{1,t-1}] + [b_{1,t-1}] \\
[m_{2t}] &= [a_{21,1} \ a_{22,1}] [z_{2t}] + [-a_{21,1} \ 1-a_{22,1}] [m_{2,t-1}] + [b_{2,t-1}]
\end{align*}
\]

(2.1.8a)

\[
\begin{align*}
[b_{1t}] &= [a_{11,2} \ a_{12,2}] [m_{1t}] + [1-a_{11,2} \ -a_{12,2}] [b_{1,t-1}] \\
[b_{2t}] &= [a_{21,2} \ a_{22,2}] [m_{2t}] + [-a_{21,2} \ 1-a_{22,2}] [b_{2,t-1}]
\end{align*}
\]

(2.1.8b)

Both the level and the slope are weighted averages. The choice of weights/smoothing matrices is not as straightforward as in simple bivariate exponential smoothing. The error-correction form of the above algorithm is
\[
\begin{bmatrix}
\mathbf{m}_{1t} \\
\mathbf{m}_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{m}_{1,t-1} \\
\mathbf{m}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{a}_{11,1} & \mathbf{a}_{12,1} \\
\mathbf{a}_{21,1} & \mathbf{a}_{22,1}
\end{bmatrix}
\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
= \begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
- \begin{bmatrix}
\mathbf{m}_{1,t-1} \\
\mathbf{m}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\mathbf{b}_{1t} \\
\mathbf{b}_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{a}_{11,2} & \mathbf{a}_{12,2} \\
\mathbf{a}_{21,2} & \mathbf{a}_{22,2}
\end{bmatrix}
\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]

Again, although not obvious from the recurrence form, the dependence between \( \mathbf{m}_{1t} \) and \( \mathbf{m}_{2t} \) (and also of \( \mathbf{b}_{1t} \) and \( \mathbf{b}_{2t} \)) is only through the errors \( \mathbf{e}_{1,t|t-1} \) and \( \mathbf{e}_{2,t|t-1} \). This is another example of a SUTSE. The above generalization suggests bivariate IMA(2,2) process. Starting from

\[
\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{m}_{1,t-1} \\
\mathbf{m}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]

\[
(I-L)\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix}
= (I-L)\begin{bmatrix}
\mathbf{m}_{1,t-1} \\
\mathbf{m}_{2,t-1}
\end{bmatrix}
+ (I-L)\begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
+ (I-L)\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]

where

\[
(I-L)\begin{bmatrix}
\mathbf{m}_{1,t-1} \\
\mathbf{m}_{2,t-1}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{b}_{1,t-2} \\
\mathbf{b}_{2,t-2}
\end{bmatrix}
+ \mathbf{A}_1 \begin{bmatrix}
\mathbf{e}_{1,t-1|t-2} \\
\mathbf{e}_{2,t-1|t-2}
\end{bmatrix}
\]

That is,

\[
(I-L)\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mathbf{b}_{1,t-1} \\
\mathbf{b}_{2,t-1}
\end{bmatrix}
+ (I + (A_1-I)L)\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]

\[
(I-L)^2\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix}
= (I-L)\begin{bmatrix}
\mathbf{b}_{1t} \\
\mathbf{b}_{2t}
\end{bmatrix}
+ (I-L)(I + (A_1-I)L)\begin{bmatrix}
\mathbf{e}_{1,t|t-1} \\
\mathbf{e}_{2,t|t-1}
\end{bmatrix}
\]
\[ = A_2 A_1 L[e_{1,t|t-1}] + (I - (2I-A_1)L + (I-A_1)L^2)[e_{1,t|t-1}]
\]
\[ e_{2,t|t-1} \]

That is,
\[
(I - 2L + L^2)[x_{1t}] = [e_{1,t|t-1}] - (2I - A_1 - A_1 A_2)[e_{1,t-1|t-2}]
\]
\[ e_{2,t-1|t-2} \]
\[ - (-I-A_1)) [e_{1,t-2|t-3}]
\]
\[ e_{2,t-2|t-3} \]

where
\[
A_1 = \begin{bmatrix}
a_{11,1} & a_{12,1} \\
2a_{11,1} & a_{22,1}
\end{bmatrix},
A_2 = \begin{bmatrix}
a_{11,2} & a_{12,2} \\
a_{21,2} & a_{22,2}
\end{bmatrix}
\]

As in previous subsections we introduce bivariate white noise
\[
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

and define the parameter matrices
\[
\begin{bmatrix}
\theta_{11,1} & \theta_{12,1} \\
\theta_{21,1} & \theta_{22,1}
\end{bmatrix} = 2I - A_1 - A_1 A_2 \tag{2.1.9}
\]
\[
\begin{bmatrix}
\theta_{11,2} & \theta_{12,2} \\
\theta_{21,2} & \theta_{22,2}
\end{bmatrix} = - (I - A_1)^2 \tag{2.1.10}
\]

then we have
\[
(I-2L+L^2)[x_{1t}] = [\varepsilon_{1t}] - \begin{bmatrix}
\theta_{11,1} & \theta_{12,1}
\end{bmatrix}\begin{bmatrix}
\varepsilon_{1,t-1}
\end{bmatrix}
\]
\[
\begin{bmatrix}
\theta_{11,2} & \theta_{12,2}
\end{bmatrix}\begin{bmatrix}
\varepsilon_{1,t-2}
\end{bmatrix}
\]
\[
(I-2L+L^2)[x_{2t}] = \begin{bmatrix}
\varepsilon_{2t}
\end{bmatrix}
\]
\[
\begin{bmatrix}
\theta_{21,1} & \theta_{22,1}
\end{bmatrix}\begin{bmatrix}
\varepsilon_{2,t-1}
\end{bmatrix}
\]
\[
\begin{bmatrix}
\theta_{21,2} & \theta_{22,2}
\end{bmatrix}\begin{bmatrix}
\varepsilon_{2,t-2}
\end{bmatrix}
\]

and it is an IMA(2,2).
Once again, invertibility and estimation of the above IMA(2,2), and (2.1.9) and (2.1.10) provide a method for determining smoothing matrices $A_1$ and $A_2$. But then with exponential smoothing we have limitations, if it is bivariate IMA(2,2) for example, we have a far more useful tool, we can apply statistical inference.

Both of the above generalisations are aimed at SUTSE. There are generalisations which are applicable to time series which influence each other, that is a bivariate time series such that $x_{1t}$ causes $z_{2t}$ and/or $z_{2t}$ causes $x_{1t}$. One such generalization is a damped trend algorithm which we discuss in the next subsection.

2.1.3 Bivariate damped trend algorithm and ARIMA(1,1,2)


We consider the recurrence form (2.1.8) with the assumption that the estimate

$$
\begin{bmatrix}
  b_{1,t-1} \\
  b_{2,t-1}
\end{bmatrix}
$$

of slope at time $t$ is the damped estimate

$$
\begin{bmatrix}
  \phi_{11} & \phi_{12} \\
  \phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
  b_{1,t-1} \\
  b_{2,t-1}
\end{bmatrix}
$$

where the distinct eigenvalues $\lambda_1$ and $\lambda_2$ of $\Phi$ are elements of the interval $(-1, 1)$. We decompose $\Phi$ using Theorem A.3 Spectral Decomposition Representation (SDR), equation (A.3.3) as

$$
\begin{bmatrix}
  \phi_{11} & \phi_{12} \\
  \phi_{21} & \phi_{22}
\end{bmatrix}
= P
\begin{bmatrix}
  \lambda_1 & 0 \\
  0 & \lambda_2
\end{bmatrix} P^{-1}
$$

Using the result stated after (A.9.3) of Lütkepohl (1991: 461) that

$$
\Phi^k = PD^k P^{-1}
$$

then

$$
\begin{bmatrix}
  \phi_{11} & \phi_{12} \\
  \phi_{21} & \phi_{22}
\end{bmatrix}^k
= P
\begin{bmatrix}
  \lambda_1^k & 0 \\
  0 & \lambda_2^k
\end{bmatrix} P^{-1}
$$
\[ \begin{bmatrix} 0 \end{bmatrix} \quad \text{as} \quad k \to \infty \]

The recurrence form is
\[
\begin{bmatrix}
    m_{1t} \\
    m_{2t}
\end{bmatrix} = A_1 \begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} + (I-A_1) \begin{bmatrix}
    m_{1,t-1} \\
    m_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
    b_{1,t-1} \\
    b_{2,t-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    b_{1t} \\
    b_{2t}
\end{bmatrix} = A_2 \begin{bmatrix}
    m_{1t} \\
    m_{2t}
\end{bmatrix} - \begin{bmatrix}
    m_{1,t-1} \\
    m_{2,t-1}
\end{bmatrix} + (I-A_2) \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
    b_{1,t-1} \\
    b_{2,t-1}
\end{bmatrix}
\]

A similar reasoning as in section 2.1.2 gives the error-correction form
\[
\begin{bmatrix}
    m_{1t} \\
    m_{2t}
\end{bmatrix} = \begin{bmatrix}
    m_{1,t-1} \\
    m_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
    b_{1,t-1} \\
    b_{2,t-1}
\end{bmatrix} + A_1 \begin{bmatrix}
    e_{1,t|t-1} \\
    e_{2,t|t-1}
\end{bmatrix}
\]

where
\[
\begin{bmatrix}
    e_{1,t|t-1} \\
    e_{2,t|t-1}
\end{bmatrix} = \begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} - \begin{bmatrix}
    m_{1,t-1} \\
    m_{2,t-1}
\end{bmatrix} - \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
    b_{1,t-1} \\
    b_{2,t-1}
\end{bmatrix}
\]

\[
\begin{bmatrix}
    b_{1t} \\
    b_{2t}
\end{bmatrix} = \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
    b_{1,t-1} \\
    b_{2,t-1}
\end{bmatrix} + A_2 A_1 \begin{bmatrix}
    e_{1,t|t-1} \\
    e_{2,t|t-1}
\end{bmatrix}
\]

We use the reasoning analogous to that in section 2.1.2 to determine the bivariate ARIMA process suggested by this algorithm.

\[
(I-\Phi L)(I-L) \begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    e_{1,t|t-1} \\
    e_{2,t|t-1}
\end{bmatrix} - (I + \Phi - A_1 - \Phi A_2 A_1) \begin{bmatrix}
    e_{1,t-1|t-2} \\
    e_{2,t-1|t-2}
\end{bmatrix} - (-\Phi(I-A_1)) \begin{bmatrix}
    e_{1,t-2|t-3} \\
    e_{2,t-2|t-3}
\end{bmatrix}
\]

We replace \( \begin{bmatrix}
    e_{1,t|t-1} \\
    e_{2,t|t-1}
\end{bmatrix} \) by white noise process \( \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix} \) and defining the
parameter matrices
\[
\Theta_1 = I + \Phi - A_1 - \Phi A_2 A_1
\]
\[
\Theta_2 = -\Phi(I - A_1)
\]

then we have (in long-hand notation)
\[
\begin{bmatrix}
1-\phi_{11}L & -\phi_{12}L \\
-\phi_{21}L & 1-\phi_{22}L
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
- \begin{bmatrix}
\theta_{11,1} & \theta_{12,2} \\
\theta_{21,1} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix}
- \begin{bmatrix}
\theta_{11,2} & \theta_{12,2} \\
\theta_{21,2} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-1} \\
\varepsilon_{2,t-1}
\end{bmatrix}
\]

which is a bivariate ARIMA(1,1,2) process.

The elements $\phi_{12}$ and $\phi_{21}$ (if they are not zero) allow $x_{1t}$ to be related to $x_{2,t-1}$ and $x_{2t}$ to $x_{1,t-1}$ and by definition of SUTSE, they are not related by error components but by the trend. That is, bivariate ARIMA(1,1,2) does not fall in the SUTSE class.

The bivariate exponential smoothing algorithms have led us to the class of bivariate ARIMA processes. The desirable properties of these processes include Granger-causality and cointegration. They are introduced and discussed in Chapters 3 and 4 respectively, and are used in the remaining parts of the dissertation.

The next section is based on the relationships between exponential smoothing algorithms and structural time series models.

### 2.2 BIVARIATE EXPONENTIAL SMOOTHING ALGORITHMS AND SUGGESTED BIVARIATE STRUCTURAL PROCESSES

We still emphasize that all (bivariate) exponential smoothing algorithms are nonstochastic. Our next focus is on the bivariate structural processes suggested by the exponential smoothing algorithms discussed in section 2.1.
2.2.1 Simple bivariate exponential smoothing algorithm suggests bivariate random walk plus noise

Harvey (1989: 19, 45)

The bivariate vector \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) is assumed to be the sum of two components

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \text{level of first element} + \text{irregular component of first element} \\ \text{level of second element} + \text{irregular component of second element} \end{bmatrix}
\]

(2.2.1)

The irregular components are at this time not assumed to be random variables.

Given an estimate \( \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix} \) (which is not random) of the level at time \( t-1 \), and the observation \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \), the simple bivariate exponential algorithm gives a method for updating \( \begin{bmatrix} m_{1,t-1} \\ m_{2,t-1} \end{bmatrix} \):

\[
\begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} = \begin{bmatrix} m_{1,t-1} + a_{11} \varepsilon_{1,t-1} + a_{12} \varepsilon_{2,t-1} \\ m_{2,t-1} + a_{21} \varepsilon_{1,t-1} + a_{22} \varepsilon_{2,t-1} \end{bmatrix}
\]

(2.2.2)

The bivariate structural model:

1. Suppose that the irregular component (2×1 vector) in (2.2.1) is \( \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \), a bivariate white noise process.

2. Next, we replace the error vector in (2.2.2) by the bivariate white noise process \( \begin{bmatrix} \gamma_{1t} \\ \gamma_{2t} \end{bmatrix} \).
3. We suppose that for all integers $t, \tau$, the white noise processes 
\[
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]
and 
\[
\begin{bmatrix}
\gamma_{1\tau} \\
\gamma_{2\tau}
\end{bmatrix}
\]
are uncorrelated.

4. Lastly before rewriting the equations, we denote the random vector of levels by 
\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
\]
Then (2.2.1) and (2.2.2) suggest the bivariate structural time series model
\[
\begin{bmatrix}
\sigma_{1t} \\
\sigma_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]
\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\gamma_{1t} \\
\gamma_{2t}
\end{bmatrix}
\]
(2.2.3)

We have that:

(i) (2.2.3) is the bivariate random walk plus noise model which is an example of bivariate structural time series models.

(ii) The bivariate random walk plus noise is a SUTSE model.

2.2.2 Bivariate generalization of Holt's exponential smoothing algorithm suggests bivariate local linear trend

Harvey (1989: 27,45)
The reasoning is analogous to that we provided in the previous section. The assumption (2.2.1) still holds, but (2.2.2) is replaced by
Together (2.2.1) and the nonstochastic algorithm (2.2.4) suggest the following structural stochastic process:

Let \( \varepsilon_{1t} \), \( \gamma_{1t} \) and \( \omega_{1t} \) be uncorrelated bivariate white noise processes.

The bivariate local linear trend is given by

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} =
\begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} =
\begin{bmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\gamma_{1t} \\
\gamma_{2t}
\end{bmatrix}
\]

This is another example of a SUTSE model.

2.2.3 Bivariate structural model suggested by the bivariate damped trend algorithm

Harvey (1989: 46, 298, 308)

Assumption (2.2.1) still holds, and the algorithm (2.2.2) is replaced by
Together, (2.2.1) and (2.2.6) suggest the following bivariate structural stochastic process:

Let \([e_{1t}, \gamma_{1t}, \omega_{1t}]\) and \([e_{2t}, \gamma_{2t}, \omega_{2t}]\) be uncorrelated bivariate white noise processes.

The bivariate damped local linear trend model is then

\[
\begin{align*}
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} & = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\gamma_{1t} \\
\gamma_{2t}
\end{bmatrix}
\end{align*}
\]  

(2.2.7a)

\[
\begin{align*}
\begin{bmatrix}
\beta_{1t} \\
\beta_{2t}
\end{bmatrix} & = \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\omega_{1t} \\
\omega_{2t}
\end{bmatrix}
\end{align*}
\]  

(2.2.7b)

The relationship is seen at (2.2.7) to be not only through the error components but through the trend component, hence this is not a SUTSE model.

The stochastic model given by (2.2.7) is the bivariate damped local linear trend process.

### 2.2.4 Remarks

All the structural models have the basic form
where \( \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \) is an unobserved component (or signal) which may take any form, and the form depends on the anticipation of the future movements one makes from the observed data, and the other component \( \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \) is a bivariate white noise random error term.

**Bivariate random walk plus noise**

This model is given by (2.2.3) and the signal component is a random walk. This model signals a level \( \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \) where there is no predictable upward or downward trend. The covariance \( \Sigma_{\gamma\gamma} \) of \( \begin{bmatrix} \gamma_{1t} \\ \gamma_{2t} \end{bmatrix} \) gives variances of each of \( \gamma_{1t} \) and \( \gamma_{2t} \) on the main diagonal, and the off-diagonal elements give correlation between them, which is a relationship between them. These off-diagonal elements explain to some extent the relationship between the components of the signal, as well as their variances. At the end the relationship between \( x_{1t} \) and \( x_{2t} \), where \( \Sigma_{\varepsilon\varepsilon} \) the covariance of \( \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \) is also involved.

If \( \Sigma_{\varepsilon\varepsilon} = O_2 \) then \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \) so that the process is a random walk.

If \( \Sigma_{\gamma\gamma} = O_2 \) then we have \( \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{bmatrix} \) a constant level so that the process (2.2.3) reduces to
because a random variable (or vector) which has a finite mean (or mean vector) and zero variance (or covariance matrix) is constant.

**Bivariate local linear trend**

The bivariate local linear trend is given by (2.2.5) where \( \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \) accounts for the current level and \( \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} \) accounts for the current slope in observed data - apart from possible relationships in the covariance matrices as mentioned for random walk plus noise.

If \( \Sigma_{\omega \omega} = \Sigma_{\gamma \gamma} = 0 \), then the slope \( \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{1, t-1} \\ \beta_{2, t-1} \end{bmatrix} \) is a constant, and the resulting process may be written

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\]

\[
\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1, t-1} \\ \mu_{2, t-1} \end{bmatrix} + \begin{bmatrix} \beta_{1} \\ \beta_{2} \end{bmatrix} + \begin{bmatrix} \gamma_{1t} \\ \gamma_{2t} \end{bmatrix}
\]

so that the process is given by a random walk with drift plus noise model.

If \( \Sigma_{\omega \omega} = \Sigma_{\gamma \gamma} = 0 \), then \( \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{1, t-1} \\ \beta_{2, t-1} \end{bmatrix} = \begin{bmatrix} \beta_{1} \\ \beta_{2} \end{bmatrix} \) and \( \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1, t-1} \\ \mu_{2, t-1} \end{bmatrix} \)

which leads to a constant level constant slope, or detrending model
where parameters \( a_1, a_2, \beta_1 \) and \( \beta_2 \) are constant. This detrending is verified by

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
a_1 \\
a_2
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} t
\]

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
a_1 \\
a_2
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} t
\]

so that we can write

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
\]

Bivariate damped local linear trend model

The model is given by (2.2.7), and \( \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} \) and \( \begin{bmatrix}
\beta_{1t} \\
\beta_{2t}
\end{bmatrix} \) are as in the usual bivariate local linear trend. The eigenvalues \( \lambda_1 \) and \( \lambda_2 \) of \( \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \)

are all less than one in absolute value (property of damping matrix) and we have also indicated using SDR that

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^k \rightarrow \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ as } k \rightarrow \infty
\]

Because the trend is damped, its contribution is less than in the local linear trend, and the limit above indicates that in the long run the trend is not reflected significantly in the process.
If $\Sigma_{\omega\omega} = O_2$ then

$$
\begin{bmatrix}
\beta_{1t}
\beta_{2t}
\end{bmatrix} = \begin{bmatrix}
\phi_{11} & \phi_{12}
\phi_{21} & \phi_{22}
\end{bmatrix}^t
\begin{bmatrix}
\beta_{1,0}
\beta_{2,0}
\end{bmatrix}
$$

and the bivariate damped local linear trend process reduces to

$$
\begin{bmatrix}
x_{1t}
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t}
\mu_{2t}
\end{bmatrix} + \begin{bmatrix}
\epsilon_{1t}
\epsilon_{2t}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\mu_{1t}
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1}
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\phi_{11} & \phi_{12}
\phi_{21} & \phi_{22}
\end{bmatrix}^t
\begin{bmatrix}
\beta_{1,0}
\beta_{2,0}
\end{bmatrix} + \begin{bmatrix}
\gamma_{1t}
\gamma_{2t}
\end{bmatrix}
$$

If we have the origin time to a very remote past (very long ago) then $t$ is sufficiently large and

$$
\begin{bmatrix}
\phi_{11} & \phi_{12}
\phi_{21} & \phi_{22}
\end{bmatrix}^t \approx \begin{bmatrix}
0 & 0
0 & 0
\end{bmatrix}
$$

and the model above is an approximate random walk plus noise model. Otherwise the model above is characterized by a random walk with damped drift plus noise model, where

$$
\begin{bmatrix}
\beta_{1,0}
\beta_{2,0}
\end{bmatrix}
$$

is the drift,

and

$$
\begin{bmatrix}
\phi_{11} & \phi_{12}
\phi_{21} & \phi_{22}
\end{bmatrix}^t
$$

is the damping matrix, and it damps more for increasing $t$. 
Again if $\Sigma_{\gamma \gamma} = \Sigma_{\omega \omega} = O_2$ and $t$ is sufficiently large, then

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^{t-1} \begin{bmatrix}
\beta_{1,0} \\
\beta_{2,0}
\end{bmatrix}
\]

\[
\approx \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix}
\]

Thus when $\Sigma_{\gamma \gamma} = \Sigma_{\omega \omega} = O_2$, in the long run the bivariate structural damped trend model approaches

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} + \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

which is similar to (2.2.8), the model obtained from random walk plus noise when the random walk is deterministic.

The damped trend model starts as model closer to the local linear trend but the trend has limited effect. With elapsing time it damps it more, and in the long run the model (because of increasing damp) is closer to the random walk plus noise.

2.3 STATE SPACE MODEL

Many time series models can be put in state space form which is considered by many authors (Lütkepohl, Harvey, Reinsel and so on) as a standardized/unified framework for analysis of time series. With this form assessment of the differences and similarities can be made. The Kalman filter is developed within this framework, and the main use of this filter is to provide forecasts and a procedure for estimating the parameters in the model. We are not going to focus on the estimation problem.
2.3.1 State space representation (SSR)


Let \( \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} \) be a bivariate time series, then the SSR of \( \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = y_t \) is given by

\[
\begin{align*}
    y_t &= Hz_t + G\varepsilon_t \\
    z_t &= Wz_{t-1} + B\varepsilon_t
\end{align*}
\]

where \( BG^T = 0 \) and \( \varepsilon_t \) is a vector white noise process, \( z_t \) is a multivariate unobserved state called state vector. Matrices \( H, G, W \) and \( B \) are called system matrices and equations (2.3.1) are called observation (or measurement) equation and state (or transition) equation respectively.

For the derivations in the forthcoming discussions we require the following assumptions:

1. \( \varepsilon_t \) be a multivariate normal random vector with \( \Sigma_{\varepsilon\varepsilon} = I \)
2. \( z_t \) and \( \varepsilon_{\tau} \) be uncorrelated for all \( t < \tau \)

Notation

\[
\begin{align*}
Y_t &= (y_1, y_2, \ldots, y_t) \\
z(t|s) &= E(z_t|Y_s) \\
\Sigma_{zy}(t|s) &= \text{cov}
\begin{bmatrix}
    z_t \\
    y_t
\end{bmatrix}
\begin{bmatrix}
    z_t \\
    y_t
\end{bmatrix}
\end{align*}
\]

\((z|y) \sim \mathcal{N}(\mu, \Sigma)\)
denotes that the conditional distribution of \( z \) given \( y \) is a multivariate normal random vector with mean vector \( \mu \) and covariance matrix \( \Sigma \).

\( \ell \) denotes the number of periods ahead for prediction (forecast horizon).

Result 2.1

(Proposition B1, Lütkepohl 1991: 480)

Let \( \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim \mathcal{N}\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \) where \( \Sigma_{22} \) is nonsingular. Then
Result 2.2
(Proposition B2, Lütkepohl 1991: 481)
Let $y: k \times 1 \sim N(\mu, \Sigma)$, and $A: m \times k$ and $c: m \times 1$ be matrix and vector of constants. Then

$$Ay + c \sim N(A\mu + c, A\Sigma A')$$

Result 2.3
(Equation (13.3.9) Lütkepohl 1991: 432)
Let $y \sim N(\mu, \Sigma_{11})$ and $z \sim N(\mu_2, \Sigma_{22})$ be independent $(k \times 1)$ random vectors. Then

$$y + z \sim N(\mu_1 + \mu_2, \Sigma_{11} + \Sigma_{22})$$

2.3.2 Kalman filter
Lütkepohl (1991: 432 - 434)
Let the SSR (2.3.1) be given and suppose that $H$, $G$, $W$, $B$ and any parameters are known. We suppose that for each natural number $t = 1, 2, ...$

$$(z_{t-1}|Y_{t-1}) \sim N(z(t-1|t-1), \Sigma_{zz}(t-1|t-1))$$

For derivation of the Kalman recursions we use induction, then for initial $t=1$ we assume that

$$(z_0|Y_0) \sim N(z(0|0), \Sigma_{zz}(0|0))$$

Now

$$\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} H & G \varepsilon_1 \\ I \end{bmatrix} \begin{bmatrix} z_1 \\ \varepsilon_1 \end{bmatrix} = \begin{bmatrix} H\varepsilon_1 + G\varepsilon_1 \\ I \end{bmatrix}$$
\[
= \begin{bmatrix} H \end{bmatrix} (Wz_0 + B\varepsilon_1) + \begin{bmatrix} G\varepsilon_1 \\
1 \\
0 \end{bmatrix} : \text{from (2.3.1b)}
\]

\[
= \begin{bmatrix} HW \end{bmatrix} z_0 + \begin{bmatrix} HB + G \end{bmatrix} \varepsilon_1 \\
W \begin{bmatrix} B \end{bmatrix}
\]

\[
E\left(\begin{bmatrix} y_1 \\
z_1 \end{bmatrix} | Y_0 \right) = \begin{bmatrix} HW \end{bmatrix} E(z_0 | Y_0) + \begin{bmatrix} HB + G \end{bmatrix} E(\varepsilon_1 | Y_0) \\
W \begin{bmatrix} B \end{bmatrix}
\]

\[
= \begin{bmatrix} HW \end{bmatrix} z(0 | 0) : E(\varepsilon_1 | Y_0) = E(\varepsilon_1 | y_0) = 0 \\
W
\]

\[
= \begin{bmatrix} HWz(0 | 0) \\
Wz(0 | 0) \end{bmatrix}
\]

and

\[
cov\left(\begin{bmatrix} y_1 \\
z_1 \end{bmatrix} | Y_0 \right) = \begin{bmatrix} HW \end{bmatrix} \text{cov}(z_0 | Y_0) \begin{bmatrix} HW \end{bmatrix}^\dagger + \begin{bmatrix} HB + G \end{bmatrix} \text{cov}(\varepsilon_1 | Y_0) \begin{bmatrix} HB + G \end{bmatrix}^\dagger \\
W \begin{bmatrix} W \\
B \end{bmatrix}
\]

\[
(z_0 \text{ and } \varepsilon_1 \text{ are uncorrelated})
\]

\[
= \begin{bmatrix} HWz_{zz}(0 | 0)W'H' & W'H' \\
W & B
\end{bmatrix} + \begin{bmatrix} HB + G \end{bmatrix} \begin{bmatrix} B'H' + G'B' \\
B & B' \end{bmatrix}
\]

\[
= \begin{bmatrix} HWz_{zz}(0 | 0)W'H' & HWz_{zz}(0 | 0)W' \\
Wz_{zz}(0 | 0)W'H' & Wz_{zz}(0 | 0)W'
\end{bmatrix}
\]

\[
+ \begin{bmatrix} HBB'H' + GG'B' & HBB'B' \\
BB'H' & BB'B' \end{bmatrix}
\]

\[
(BG' = 0, GB' = 0)
\]
That is, \[
\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} | Y_0 \] is distributed multivariate normal with mean vector \[
\begin{bmatrix} \mu_{y(l I O)} \\ \mu_{z(l I O)} \end{bmatrix}
\] and covariance matrix \[
\begin{bmatrix} \Sigma_{yy}(l I O) & \Sigma_{yz}(l I O) \\ \Sigma_{zy}(l I O) & \Sigma_{zz}(l I O) \end{bmatrix}
\]

\[
\begin{bmatrix} y(1|0) \\ z(1|0) \end{bmatrix} = \begin{bmatrix} HWz(0|0) \\ Wz(0|0) \end{bmatrix}
\]

\[
= \begin{bmatrix} HW\Sigma_{zz}(0|0)W'H' + HBB'H' + GG' & HW\Sigma_{zz}(0|0)W'H' + HBB' \\ W\Sigma_{zz}(0|0)W'H' + BB'H' & W\Sigma_{zz}(0|0)W' + BB' \end{bmatrix}
\]

An application of result 2.1 gives

\[
(z_1|Y_1) \sim N(\mu_{z(1|1)}, \Sigma_{zz}(1|1))
\]

where

\[
z(1|1) = z(1|0) + \Sigma_{zy}(1|0)\Sigma_{yy}^{-1}(1|0)(y_1 - y(1|0))
\]

and

\[
\Sigma_{zz}(1|1) = \Sigma_{zz}(1|0) - \Sigma_{zy}(1|0)\Sigma_{yy}^{-1}(1|0)\Sigma_{yz}(1|0).
\]

Let \( t \in \mathbb{N}, \ t > 1 \) and suppose

\[
(z_{t-1}|Y_{t-1}) \sim N(\mu_{z(t-1|t-1)}, \Sigma_{zz}(t-1|t-1))
\]

then, reasoning which is analogous to that used when \( t=1 \) shows that

\[
\begin{bmatrix} y_t \\ z_t \end{bmatrix} | Y_{t-1} \sim N\left(\begin{bmatrix} y(t|t-1) \\ z(t|t-1) \end{bmatrix}, \begin{bmatrix} \Sigma_{yy}(t|t-1) & \Sigma_{yz}(t|t-1) \\ \Sigma_{zy}(t|t-1) & \Sigma_{zz}(t|t-1) \end{bmatrix}\right)
\]

where

\[
\begin{bmatrix} y(t|t-1) \\ z(t|t-1) \end{bmatrix} = \begin{bmatrix} HWz(t-1|t-1) \\ Wz(t-1|t-1) \end{bmatrix}
\]

\[\text{(2.3.2)}\]

and
\[
\begin{bmatrix}
\Sigma_{yy}(t|t-1) & \Sigma_{yz}(t|t-1) \\
\Sigma_{zy}(t|t-1) & \Sigma_{zz}(t|t-1)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
HW\Sigma_{zz}(t-1|t-1)W' + HBB' H' + GG' \\
W\Sigma_{zz}(t-1|t-1)W' + BB' H' 
\end{bmatrix}
\]

Finally, by employing Result 2.1 we obtain

\[
(z_t | Y_t) \sim N(z(t|t), \Sigma_{zz}(t|t))
\]

where

\[
z(t|t) = z(t|t-1) + \Sigma_{zy}(t|t-1)\Sigma_{yy}^{-1}(t|t-1)(y_t - y(t|t-1))
\]

and

\[
\Sigma_{zz}(t|t) = \Sigma_{zz}(t|t-1) - \Sigma_{zy}(t|t-1)\Sigma_{yy}^{-1}(t|t-1)\Sigma_{zy}(t|t-1)
\]

### 2.3.3 Kalman forecasts

Lütkepohl (1991: 432 - 434)

We let \( t \) be our forecast origin and \( \ell \) a forecast horizon, that is we intend forecasting \( \ell \) periods ahead. Suppose that we are given

\[
(z_t | Y_t) \sim N(z(t|t), \Sigma_{zz}(t|t))
\]

Let \( \ell = 1 \), then using the Kalman filter (2.3.2) we replace \( t \) by \( t+1 \) and we obtain

\[
\begin{bmatrix}
y(t+1|t) \\
z(t+1|t)
\end{bmatrix} = 
\begin{bmatrix}
HWz(t|t) \\
Wz(t|t)
\end{bmatrix}
\]

and using (2.3.3) we obtain

\[
\Sigma_{zz}(t+1|t) = W\Sigma_{zz}(t|t)W' + BB'
\]

(2.3.5a)

\[
\Sigma_{yy}(t+1|t) = H\Sigma_{zz}(t+1|t)H' + GG'
\]

(2.3.5b)
We now want to derive similar expressions for $\ell = 2, 3, ...$

Result 2.4

Suppose we are given that

$$z(t+\ell-1|Y_t) \sim N(z(t+\ell-1|t), \Sigma_{zz}(t+\ell-1|t))$$

Then

$$\begin{bmatrix}
y_{t+\ell} \\
z_{t+\ell}
\end{bmatrix}$$

is distributed multivariate normal with mean vector

$$\begin{bmatrix}
y(t+\ell|t) \\
z(t+\ell|t)
\end{bmatrix} = \begin{bmatrix}
HWz(t+\ell-1|t) \\
Wz(t+\ell-1|t)
\end{bmatrix}$$

and covariance matrix

$$\begin{bmatrix}
\Sigma_{yy}(t+\ell|t) & \Sigma_{yz}(t+\ell|t) \\
\Sigma_{zy}(t+\ell|t) & \Sigma_{zz}(t+\ell|t)
\end{bmatrix} = \begin{bmatrix}
H\Sigma_{zz}(t+\ell|t)H' + GG' & H\Sigma_{zz}(t+\ell|t) \\
\Sigma_{zz}(t+\ell|t)H' & W\Sigma_{zz}(t+\ell-1|t)W' + BB'
\end{bmatrix}$$

Proof

We use induction on $\ell$ to prove this result.

The proof of the Kalman recursions shows that the result is true when $\ell = 1$.

Suppose that the result holds for $k = \ell - 1 \in \mathbb{N}, \quad \ell \geq 2$. That is we assume that

$$z(t+\ell-1|Y_t) \sim N(z(t+\ell-1|t), \Sigma_{zz}(t+\ell-1|t)) \quad (2.3.6a)$$

implies that

$$\begin{bmatrix}
y(t+\ell-1) \\
z(t+\ell-1)
\end{bmatrix}$$
is distributed multivariate normal with mean vector

$$\begin{bmatrix}
y(t+\ell-1) \\
z(t+\ell-1)
\end{bmatrix}$$
\[
\begin{align*}
[y(t+\ell-1|t)] &= [HWz(t+\ell-2|t)] \\
z(t+\ell-1|t) &= [Hz(t+\ell-2|t)] \\
\end{align*}
\] (2.3.6b)

and covariance matrix

\[
\begin{bmatrix}
\Sigma_{yy}(t+\ell-1|t) & \Sigma_{yz}(t+\ell-1|t) \\
\Sigma_{zy}(t+\ell-1|t) & \Sigma_{zz}(t+\ell-1|t)
\end{bmatrix}
= \begin{bmatrix}
[HW^\Sigma_{zz}(t+\ell-1|t)H^I + GG^I & H\Sigma_{zz}(t+\ell-1|t)W^I + BB^I] \\
\Sigma_{zz}(t+\ell-1|t)H^I & W\Sigma_{zz}(t+\ell-1|t)W^I + BB^I
\end{bmatrix}
\] (2.3.6c)

We show that if the above case of \( \ell-1 \) is true, then \( \ell \) will also satisfy the recursions. With the similar approach we have been following we write

\[
\begin{align*}
[y(t+\ell)] &= [HW]z(t+\ell-1) + [HB+G]\varepsilon(t+\ell) \\
z(t+\ell) &= [W]
\end{align*}
\]

then

\[
\begin{align*}
[y(t+\ell|t)] &= [HW]z(t+\ell-1|t) \\
z(t+\ell|t) &= [W]
\end{align*}
\] (2.3.7)

and

\[
\begin{align*}
\begin{bmatrix}
\Sigma_{yy}(t+\ell|t) & \Sigma_{yz}(t+\ell|t) \\
\Sigma_{zy}(t+\ell|t) & \Sigma_{zz}(t+\ell|t)
\end{bmatrix}
= \begin{bmatrix}
[HW\Sigma_{zz}(t+\ell-1|t)H^I + GG^I] & H\Sigma_{yz}(t+\ell|t) \\
\Sigma_{zz}(t+\ell|t)H^I & W\Sigma_{zz}(t+\ell-1|t)W^I + BB^I
\end{bmatrix}
\] (2.3.8)

and this completes the proof. \( \blacksquare \)
2.3.4 Formulae for point forecasts

The Kalman forecasts gave all the formulae we require in this section, we formalize them and make them convenient for use in point forecasts for statistical time series models.

(Lütkepohl 1991: 432 - 434)

Suppose we are given that

\[ (z_t \mid Y_t) \sim N(z(t \mid t), \Sigma_{zz}(t \mid t)) \]

then (2.3.4) gives for \( \ell = 1 \)

\[
\begin{bmatrix}
 y(t+1 \mid t) \\
 z(t+1 \mid t)
\end{bmatrix} =
\begin{bmatrix}
 HWz(t \mid t) \\
 Wz(t \mid t)
\end{bmatrix}
\] (2.3.9)

and for \( \ell = 2 \) gives

\[
\begin{bmatrix}
 y(t+2 \mid t) \\
 z(t+2 \mid t)
\end{bmatrix} =
\begin{bmatrix}
 HWz(t+1 \mid t) \\
 Wz(t+1 \mid t)
\end{bmatrix}
\] (2.3.10)

From (2.3.9) we deduce (amongst others) that

\[ z(t+1 \mid t) = Wz(t \mid t), \] (2.3.11)

which when substituted in (2.3.11) we obtain

\[
\begin{bmatrix}
 y(t+2 \mid t) \\
 z(t+2 \mid t)
\end{bmatrix} =
\begin{bmatrix}
 HW^2z(t \mid t) \\
 W^2z(t \mid t)
\end{bmatrix}
\] (2.3.12)

The induction step (2.3.6) becomes that we assume that \( \ell-1 \) results in the same pattern as above, namely

\[
\begin{bmatrix}
 y(t+\ell-1) \\
 z(t+\ell-1)
\end{bmatrix} =
\begin{bmatrix}
 HW^{\ell-1}z(t \mid t) \\
 W^{\ell-1}z(t \mid t)
\end{bmatrix}
\]
and this implies that

\[ z(t+\ell-1|t) = W^{\ell-1}z(t|t) \]  

(2.3.13)

From (2.3.7) for case \( \ell \) periods ahead

\[ \begin{bmatrix} y(t+\ell|t) \\ z(t+\ell|t) \end{bmatrix} = \begin{bmatrix} HW \\ W \end{bmatrix} z(t+\ell-1|t) \]

\[ = \begin{bmatrix} HW \\ W \end{bmatrix} W^{\ell-1}z(t|t) \]

: using (2.3.13)

\[ = \begin{bmatrix} HW^\ell z(t|t) \\ W^\ell z(t|t) \end{bmatrix} \]

That is

\[ y(t+\ell|t) = HW^\ell z(t|t) \]  

(2.3.14)

2.3.5 Formulae for forecast MSE

Lütkepohl (1991: 432 - 434)

This is again formalization of subsection 2.3.3 and restatement of covariance matrices (herein referred to as MSE for forecasts) in a more convenient form.

Suppose that we are given \( \Sigma_{zz}(t|t) \) and we want the MSE's for forecast horizons \( \ell \).

As before the derivation is done by induction on \( \ell \).

\( \ell = 1 \)

Writing the SSR (2.3.1) with \( t \) replaced by \( t+1 \) and a more convenient form

\[ \begin{bmatrix} y(t+1) \\ z(t+1) \end{bmatrix} = \begin{bmatrix} HW \\ W \end{bmatrix} z(t) + \begin{bmatrix} HB+G \\ B \end{bmatrix} \epsilon(t+1) \]
and the assumptions given, then

\[
\begin{bmatrix}
\Sigma_{yy}(t+1|t) & \Sigma_{yz}(t+1|t) \\
\Sigma_{zy}(t+1|t) & \Sigma_{zz}(t+1|t)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
HW\Sigma_{zz}(t|t)W'W & HW\Sigma_{zz}(t|t)W'W \\
W\Sigma_{zz}(t|t)W'W & W\Sigma_{zz}(t|t)W'W
\end{bmatrix}
+ \begin{bmatrix}
HB'B'+GG' & HBB' \\
BB'H' & BB'
\end{bmatrix}
\]

\[
= \begin{bmatrix}
HW^2\Sigma_{zz}(t|t)W'(W)^2 & HWBB'W'W \\
W\Sigma_{zz}(t|t)W'W & W\Sigma_{zz}(t|t)(W')^2+WBB'W'
\end{bmatrix}
+ \begin{bmatrix}
HB'B'+GG' & HBB' \\
BB'H' & BB'
\end{bmatrix}
\]

\[
(t=2)
\]

The SSR for \( t \) replaced by \( t+2 \) is

\[
\begin{bmatrix}
y(t+2) \\
z(t+2)
\end{bmatrix}
= \begin{bmatrix}
HW & z(t+1) \\
W & B
\end{bmatrix}
\]

Using the assumptions \( BG' = 0, GB' = 0, z(t) \) and \( \epsilon(t+1) \) are uncorrelated the MSE becomes:

\[
\begin{bmatrix}
\Sigma_{yy}(t+2|t) & \Sigma_{yz}(t+2|t) \\
\Sigma_{zy}(t+2|t) & \Sigma_{zz}(t+2|t)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
HW\Sigma_{zz}(t+1|t)W'W & HW\Sigma_{zz}(t+1|t)W'W \\
W\Sigma_{zz}(t+1|t)W'W & W\Sigma_{zz}(t+1|t)W'W
\end{bmatrix}
+ \begin{bmatrix}
HB'B'+GG' & HBB' \\
BB'H' & BB'
\end{bmatrix}
\]

\[
= \begin{bmatrix}
HW^2\Sigma_{zz}(t|t)(W')^2 & HWBB'W'W \\
W\Sigma_{zz}(t+1|t)W'W & W\Sigma_{zz}(t+1|t)(W')^2+WBB'W'
\end{bmatrix}
+ \begin{bmatrix}
HB'B'+GG' & HBB' \\
BB'H' & BB'
\end{bmatrix}
\]
That is, for $\ell = 1$
\[
    \Sigma_{zz}(t+1|t) = W\Sigma_{zz}(t|t)W^t + BB'
\]
\[
    \Sigma_{yy}(t+1|t) = H\Sigma_{zz}(t+1|t)H^t + GG'
\]
\[
= HW\Sigma_{zz}(t|t)W'H^t + HBB'H^t + GG'
\]
and for $\ell = 2$
\[
    \Sigma_{zz}(t+2|t) = W^2\Sigma_{zz}(t|t)(W')^2 + WBB'W' + BB'
\]
\[
    \Sigma_{yy}(t+2|t) = HW^2\Sigma_{zz}(t|t)(W')^2H' + HWBB'W'H' + HBB'H' + GG'
\]

The induction step is:
we assume that for $\ell - 1$ (where $W^0 = I$)
\[
    \Sigma_{zz}(t+\ell-1|t) = W^{\ell-1}\Sigma_{zz}(t|t)(W')^{\ell-1} + \sum_{j=0}^{\ell-2} W^jBB'(W')^j
\]
\[
    \Sigma_{yy}(t+\ell-1|t) = HW^{\ell-1}\Sigma_{zz}(t|t)(W')^{\ell-1}H' + \sum_{j=0}^{\ell-2} HW^jBB'(W')^jH' + GG'
\]

For the case $\ell$-steps we use (2.3.8)
\[
    \Sigma_{zz}(t+\ell|t) = W\Sigma_{zz}(t+\ell-1|t)W' + BB'
\]
\[
    \Sigma_{yy}(t+\ell|t) = H\Sigma_{zz}(t+\ell|t)H' + GG'
\]

Using induction step for $\Sigma_{zz}(t+\ell|t)$ we obtain
\[
    \Sigma_{zz}(t+\ell|t) = W^\ell \Sigma_{zz}(t|t)(W')^\ell + \sum_{j=0}^{\ell-2} W^{j+1}BB'(W')^{j+1} + BB'
\]
\[
= W^\ell \Sigma_{zz}(t|t)(W')^\ell + \sum_{j=0}^{\ell-1} W^{j}BB'(W')^{j} \tag{2.3.15}
\]

and substituting this for $\Sigma_{yy}(t+\ell|t)$ we have
\[
    \Sigma_{yy}(t+\ell|t) = H[W^\ell \Sigma_{zz}(t|t)(W')^\ell + \sum_{j=0}^{\ell-1} W^{j}BB'(W')^{j}]H' + GG'
\]

That is
\[
    \Sigma_{yy}(t+\ell|t) = HW^\ell \Sigma_{zz}(t|t)(W')^\ell + \sum_{j=0}^{\ell-1} HW^{j}BB'(W')^{j}H' + GG' \tag{2.3.16}
\]
CHAPTER 3

CAUSALITY (AND FEEDBACK)
OF ARIMA MODELS

When considering a bivariate time series with components $x_{1t}$ and $x_{2t}$, in addition to the properties of univariate time series for each component, we may find that amongst others, one time series influences/causes the other, or they influence each other. We study the conditions for existence of these relationships, and the three types discussed are: Granger-causality, instantaneous causality and feedback. The bivariate processes we focus on are assumed to have zero mean vector.

3.1 DEFINITIONS AND ASSUMPTIONS

We need to remember the definition given in the previous chapter of bivariate white noise process.

Definition (ARIMA(p,d,q))

Tsay & Tiao (1990: 220)

A bivariate process $\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$ is said to be autoregressive-integrated moving average of orders p, d and q (ARIMA(p,d,q)) if it can be expressed in the form

$$V = 1 - L$$

$$\phi_{ij}(L) = \delta_{ij} - \phi_{ij,1}L - \phi_{ij,2}L^2 - \ldots - \phi_{ij,p}L^p$$

$$\theta_{ij}(L) = \delta_{ij} + \theta_{ij,1}L + \theta_{ij,2}L^2 + \ldots + \theta_{ij,q}L^q$$

$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{bmatrix} v^d x_{1t} \\ v^d x_{2t} \end{bmatrix} = \begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

(3.1.1)
\[ \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \]

and

\[ i = 1, 2; \quad j = 1, 2 \]

In ARIMA\((p,d,q)\) there is an autoregressive part of order \(p\), AR\((p)\), integration of order \(d\), I\((d)\) (this concept will be explained after the next definition) and the moving average part of order \(q\), MA\((q)\). The definition of stationarity which follows, helps in the understanding of I\((d)\).

**Definition (Stationary process)**

Lütkepohl (1991: 19), Reinsel (1993: 2)

A stochastic process \([x_{1t} \ x_{2t}]\) is said to be stationary if it satisfies

\[(1) \ E \left[ \begin{array}{c} x_{1t} \\ x_{2t} \end{array} \right] = \left[ \begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right] \quad \text{for all } t \quad (3.1.2a)\]

\[(2) \ E \left[ \begin{array}{c} x_{1t} - \mu_1 \\ x_{2t} - \mu_2 \end{array} \right] \left[ \begin{array}{c} x_{1t-h} - \mu_1 \\ x_{2t-h} - \mu_2 \end{array} \right]^T = \Gamma(h) = \Gamma^T(-h) \quad \text{for all } t, h \geq 0 \quad (3.1.2b)\]


Accordingly, a process is stationary if it is constant (stationary) in first and second moments. If they are not stationary they would vary with time and thereby written \(\mu_t\) and \(\Gamma_t(h)\) for the time period \(t\). When only one of these moments is not stationary, say \(\mu_t\), we say that the process is nonstationary in mean, and if the covariance matrix is not stationary we speak of nonstationary in variances. Nonstationarity therefore, may be in one of them, or in both. We have assumed that the processes we are considering have mean vector \(\mu_t = 0\), and this leads to that whenever we seek to verify nonstationarity (or stationarity), only the covariance should be considered.
There are time series processes which are nonstationary, but are thought to "approach" or become stationary after being differenced a (specific) number of times. We define them.

Definition (Integrated process)
A univariate nonstationary time series process is said to be integrated of order d, I(d), if it requires to be differenced d times before it becomes stationary.

One property is that $y_t$ is integrated of order d ($y_t \sim I(d)$), then
$$v^k y_t$$ is nonstationary for $k = 0, 1, \ldots, d-1$
and
$$v^d y_t$$ is stationary.

Also, one way of saying that $x_t$ is stationary is to say $x_t \sim I(0)$, that is, it requires no differencing to make it stationary as it is already stationary. An ARIMA$(p,d,q) \sim I(d)$ and when $d=0$, then we have ARMA$(p,q)$, and of course it is stationary.

If a process is integrated (of order d, say) we note that the moments can be stationary only after differencing (d times), otherwise they vary over time. Nonstationary processes have additional properties which are not found in stationary ones, for example cointegration may exist in a nonstationary process, but will never occur in a stationary model. Cointegration of ARMA processes will be discussed in Chapter 4.

The AR matrix $\Phi(L)$ of parameters $\phi_{ij}$ in ARIMA may be written as
$$\begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} = [1 \ 0] - \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} - \ldots - \begin{bmatrix} \phi_{11,p}L^p & \phi_{12,p}L^p \\ \phi_{21,p}L^p & \phi_{22,p}L^p \end{bmatrix}$$
and the MA one $\Theta(L)$ of $\theta_{ij}$ as
$$\begin{bmatrix} \theta_{11}(L) & \theta_{12}(L) \\ \theta_{21}(L) & \theta_{22}(L) \end{bmatrix} = [1 \ 0] + \begin{bmatrix} \theta_{11,1}L & \theta_{12,1}L \\ \theta_{21,1}L & \theta_{22,1}L \end{bmatrix} + \ldots + \begin{bmatrix} \theta_{11,q}L^q & \theta_{12,q}L^q \\ \theta_{21,q}L^q & \theta_{22,q}L^q \end{bmatrix}$$
From the discussions in Harvey (1989: 64-65), Lütkepohl (1991: 10-11) and Reinsel (1993: 12-13), especially by noting use of decomposition SDR using eigenvalues, we realize that the eigenvalues of the AR and MA matrices are of importance in the properties of time series. One of the convenient approaches in studying these characteristics is to use the so-called fundamental MA or prediction error MA representation which is defined on stable processes (defined shortly), and we start with a simple case, AR(1).

If in (3.1.1) we have \( p=1, \ d=q=0 \) then we have a bivariate AR(1) process which is

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix}
\tag{3.1.3}
\]

where \( \begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix} \) is a bivariate white noise process.

**Definition (Stable bivariate AR(1))**


A bivariate AR(1) process (3.1.3) is said to be stable if

\[
\det \begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix}
- \begin{bmatrix}
    \phi_{11} & \phi_{12} \\
    \phi_{21} & \phi_{22}
\end{bmatrix}z \neq 0 \text{ for } |z| \leq 1
\tag{3.1.4}
\]

We call (3.1.5) a stability condition, and it simply means that the determinant

\[
\begin{vmatrix}
    1-\phi_{11}z & -\phi_{12}z \\
    -\phi_{21}z & 1-\phi_{22}z
\end{vmatrix}
\]

does not have zeros inside or on the unit circle. Lütkepohl (1991: 10) and Reinsel (1993: 26) have it using eigenvalues as that this condition, (3.1.4), implies that all eigenvalues of \( \Phi \), the AR matrix, are less than one in absolute value.
When the stability condition holds

\[
\begin{bmatrix}
1 - \phi_{11} L & -\phi_{12} L \\
-\phi_{21} L & 1 - \phi_{22} L
\end{bmatrix}^{-1}
= \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
+ \sum_{j=1}^{\infty} \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^j L^j.
\]

A stable bivariate AR(1) process can therefore be written as the infinite order moving average process

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
+ \sum_{j=1}^{\infty} \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^j \begin{bmatrix}
\varepsilon_{1,t-j} \\
\varepsilon_{2,t-j}
\end{bmatrix}.
\]

This is the fundamental MA or prediction error MA representation of the stable bivariate AR(1) process.

The fundamental representation we have mentioned earlier is a special MA representation, very important in that it is the basis of the proof of the Granger-causality criterion (inter alia).

We develop this representation on a general AR process, from (3.1.1) with d=q=0, the result is an AR(p) process

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\phi_{11,1} & \phi_{12,1} \\
\phi_{21,1} & \phi_{22,1}
\end{bmatrix} \begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix}
+ \cdots + \begin{bmatrix}
\phi_{11,p} & \phi_{12,p} \\
\phi_{21,p} & \phi_{22,p}
\end{bmatrix} \begin{bmatrix}
x_{1,t-p} \\
x_{2,t-p}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

(3.1.5)

To establish the stability condition on (3.1.5) we write (3.1.5) as

\[
x_t = \Phi_1 x_{t-1} + \cdots + \Phi_p x_{t-p} + \varepsilon_t
\]

For SSR, let

\[
\mathbf{x}_t = \begin{bmatrix}
\mathbf{x}_t \\
\mathbf{x}_{t-1} \\
\vdots \\
\mathbf{x}_{t-p+1}
\end{bmatrix}
\]

(2p×1)

\[
= \begin{bmatrix}
x_t \\
x_{t-1} \\
\vdots \\
x_{t-p+1}
\end{bmatrix}
\]
then
\[
\begin{bmatrix}
\Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + \varepsilon_t \\
x_{t-1} \\
x_{t-2} \\
\vdots \\
x_{t-p+1}
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\Phi_1 & \Phi_2 & \cdots & \Phi_p \\
I_2 & O_2 & \cdots & O_2 \\
O_2 & I_2 & \cdots & O_2 \\
\vdots & \vdots & \ddots & \vdots \\
O_2 & O_2 & \cdots & I_2
\end{bmatrix}
\begin{bmatrix}
z_{t-1} \\
I_2 \\
O_2 \\
\vdots \\
O_2
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_t \\
O_2 \\
O_2 \\
\vdots \\
O_2
\end{bmatrix}
\]

and
\[
x_t = \begin{bmatrix}
I_2 & O_2 & \cdots & O_2
\end{bmatrix} z_t
\]

Let
\[
W = \begin{bmatrix}
\Phi_1 & \Phi_2 & \cdots & \Phi_p \\
I_2 & O_2 & \cdots & O_2 \\
O_2 & I_2 & \cdots & O_2 \\
\vdots & \vdots & \ddots & \vdots \\
O_2 & O_2 & \cdots & I_2
\end{bmatrix},
B = \begin{bmatrix}
I_2 \\
O_2 \\
O_2 \\
\vdots \\
O_2
\end{bmatrix},
H' = \begin{bmatrix}
I_2 \\
O_2 \\
O_2 \\
\vdots \\
O_2
\end{bmatrix}
\]

then
\[
z_t = Wz_{t-1} + Be_t 
\tag{3.1.6a}
\]
\[
x_t = Hz_t 
\tag{3.1.6b}
\]

The stability condition (3.1.4) can be extended to \((2p \times 1)\) AR(1) process \(z_t\) as

\[
|I_{2p} - Wz| \neq 0 \text{ if } |z| \leq 1
\tag{3.1.7}
\]
Now
\[ |I_{2p} - Wz| = \begin{vmatrix}
I_2 - \Phi_1 z & -\Phi_2 z & \cdots & -\Phi_p z \\
-I_2 z & I_2 & \cdots & 0 \\
0 & -I_2 z & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & -I_2 z & I_2 \\
\end{vmatrix} \]

To derive the above determinant we perform column operations as follows:
- multiply (2p)th column by \( z \) and add to (2p-1)st column
- multiply (2p-1)st column by \( z \) and add to (2p-2)nd column
- multiply second column by \( z \) and add to first column
then
\[ |I_{2p} - Wz| = \begin{vmatrix}
\Phi(z) & \Phi(1)(z) & \Phi(2)(z) & \cdots & \Phi(p-1)(z) \\
O_2 & I_2 & O_2 & \cdots & O_2 \\
O_2 & O_2 & I_2 & \cdots & O_2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
O_2 & O_2 & O_2 & \cdots & I_2 \\
\end{vmatrix} \]

which is the determinant of a triangular matrix of diagonal elements
\[ \Phi(z) = I_2 - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p \]

and the \( I_2 \).

The other nonzero elements are given by
\[ \Phi(1)(z) = -\Phi_2 z - \Phi_3 z^2 - \cdots - \Phi_p z^{p-1} \]
\[ \Phi(2)(z) = -\Phi_3 z - \Phi_4 z^2 - \cdots - \Phi_p z^{p-2} \]
\[ \vdots \]
\[ \Phi(p-2)(z) = -\Phi_{p-1} z - \Phi_p z^2 \]
\[ \Phi(p-1)(z) = -\Phi_p z \]
Using Lütkepohl (1991: 452 Rule (3)) the determinant of a triangular matrix is given by product of diagonal elements, hence we obtain

\[ |I_{2p} - Wz| = |I_2 - \Phi_1 z - \ldots - \Phi_p z^p| \]

so that the stability condition (3.1.7) becomes

\[ |I_2 - \Phi_1 z - \ldots - \Phi_p z^p| \neq 0 \text{ if } |z| < 1 \] (3.1.8)

Developing (3.1.6a) further we have

\[
z_t = Wz_{t-1} + \gamma_t, \quad \gamma_t = B\varepsilon_t \\
= W(Wz_{t-2} + \gamma_{t-1}) + \gamma_t \\
= W^2 z_{t-2} + W\gamma_{t-1} + \gamma_t
\]

Repeating the process we obtain

\[
z_t = \sum_{i=0}^{\infty} W^i \gamma_{t-i}, \quad W^0 = I_{2p}
\] (3.1.9)

Using (3.1.6b), where \( H = \begin{bmatrix} I_2 & 0 & \cdots & 0 \end{bmatrix} \)

\[
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = Hz_t \\
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = \sum_{i=0}^{\infty} HW^i \gamma_{t-i}
\]

Now

\[
H^t H \gamma_t = \begin{bmatrix} I_2 & I_2 & \cdots & I_2 \\ I_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ I_2 & O_2 & \cdots & O_2 \end{bmatrix} \gamma_t
\]
\[ H'HisHt = \begin{bmatrix} I_2 \\ O_2 \\ \vdots \\ O_2 \end{bmatrix} \varepsilon_t, \quad \text{since} \quad \gamma_t = \begin{bmatrix} I_2 \\ O_2 \\ \vdots \\ O_2 \end{bmatrix} \varepsilon_t \]

That is

\[ \gamma_t = \begin{bmatrix} \gamma_t \\ \gamma_t \end{bmatrix} \]

\[ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\ \psi_{21,i} & \psi_{22,i} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix} : \quad \gamma_{t-i} = H'HisH_{t-i} \]

where

\[ \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\ \psi_{21,i} & \psi_{22,i} \end{bmatrix} = HW^iH' \]

and

\[ H\gamma_t = \begin{bmatrix} I_2 & O_2 & \ldots & O_2 \end{bmatrix} \begin{bmatrix} I_2 \\ O_2 \\ \vdots \\ O_2 \end{bmatrix} \varepsilon_{1t} \]

\[ = I_2 \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \]


The stability condition (3.1.7) is equivalent to that all eigenvalues \( \lambda_k \) of \( W \) are less than or equal to one in absolute value. Also, the eigenvalues of \( W_i \) are \( \lambda_k^i \) (cf Graybill 1983: 44 Theorem 3.2.6) which are also less than or equal to one in absolute value.
Also
\[
\begin{vmatrix}
\psi_{11,1} & \psi_{12,1} \\
\psi_{21,1} & \psi_{22,1}
\end{vmatrix}
= |HH^iH^i|
\]
\[
= |HH^iW^i| : |AB| = |BA| \text{ if multiplication is defined}
\]
\[
= |W^i|
\]

which shows that eigenvalues of matrix $\Psi_i$ and of $W^i$ are the same. (cf Rule 5 Lütkepohl 1991: 456.) That is, the eigenvalues of $\Psi_i$ are $\lambda_k^i$, all less than or equal to one in absolute value.

The manipulation from (3.1.9) up to (3.1.10) can be more easily understood in an easy case such as $p=1$. The following approach was suggested by Prof. Markham. Starting from (3.1.9) we have

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\gamma_{1t} \\
\gamma_{2t}
\end{bmatrix}
+ \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
\gamma_{1,t-1} \\
\gamma_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\phi_{11} & \phi_{12}
\end{bmatrix}^2
\begin{bmatrix}
\gamma_{1,t-2} \\
\gamma_{2,t-2}
\end{bmatrix}
+ \cdots
\]

where
\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^2
= \begin{bmatrix}
\phi_{11}^2 + \phi_{12}^2 & \phi_{11}\phi_{12} + \phi_{12}\phi_{22} \\
\phi_{21}\phi_{11} + \phi_{22}\phi_{21} & \phi_{21}\phi_{12} + \phi_{22}^2
\end{bmatrix}
\]

and so on.

Then
\[
x_{1t} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}
\]

\[
= \sum_{i=0}^{\infty} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^i \begin{bmatrix} \gamma_{1,t-i} \\
\gamma_{2,t-i}
\end{bmatrix} 
\]

(3.1.11)

\[
= \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\
\varepsilon_{2,t-i}
\end{bmatrix}
\]

(3.1.12)

where
\[
\begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^i
\]
Theorem 3.1
A stable bivariate AR(p) process is stationary

Proof
Let \[
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix}
\]
be a stable AR(p) process as in (3.1.5). Then it has the SSR given by (3.1.9) whose state vector is
\[
x_t = \sum_{i=0}^{n} W^i \gamma_{t-i}, \quad W^0 = I_{2p}
\]
We have
\[
E(\gamma_t) = \begin{bmatrix} I_2 \end{bmatrix} E(\varepsilon_t)
\]
\[
= 0: 2p \times 1
\]
That is
\[
E(x_t) = 0: 2p \times 1
\]
\[
E(\gamma_{t-h-i} \gamma_{t-i}) = \begin{bmatrix} I_2 \end{bmatrix} E(\varepsilon_{t-h-i} \varepsilon_{t-i}^t) [I_2 \quad O_2 \ldots O_2]
\]
Now
\[
E(\varepsilon_{t-h-i} \varepsilon_{t-i}^t) = E\left[\begin{bmatrix}
  \varepsilon_{1,t-h-i} \\
  \varepsilon_{2,t-h-i}
\end{bmatrix} \begin{bmatrix}
  \varepsilon_{1,t-i} & \varepsilon_{2,t-i}
\end{bmatrix}\right]
\]
The autocovariance of \( z_t \) is given by

\[
\Gamma_t(h) = E(z_t z_{t-h})
\]

Then

\[
E(\gamma_{t-h-i}^t \gamma_{t-i}^t) = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{c} 0 \\ 0 \end{array} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

where

\[
\Sigma = \begin{bmatrix} \Sigma_{\varepsilon \varepsilon} & 0 \\ 0 & 0 \end{bmatrix}
\]

From the measurement equation (3.1.6b) we have

\[
z_t = [I_2 \ 0_2 ... 0_2] z_t
\]

Then

\[
E(z_t) = [I_2 \ 0_2 ... 0_2] E(z_t) = 0
\]

The autocovariance matrix is

\[
\Gamma_{z,t}(h) = E(H z_t z_{t-h}^H)
\]
Both the mean vector and the autocovariance matrix of are time invariant and therefore we conclude that it is stationary.

### 3.2 More Definitions: Forecasting Concepts


One of the main objectives of time series analysis is forecasting. In many instances a forecaster is required in a particular period \( t \) to make statements about values \( x_1, x_2, \ldots, x_t \) for time periods \( t+\ell, \ell = 1, 2, \ldots \). The requirements for this task are a model to generate data, and a set \( U_t \) containing relevant information in period \( t \). For our bivariate case we define

\[
U_t = \left\{ \begin{bmatrix} x_{1s} \\ x_{2s} \end{bmatrix} \mid s \leq t \right\} \tag{3.2.1}
\]

The set \( U_t \) contains past and present information (at time \( t \)) about the process \( x_t \). There are cases where \( s < t \) is used, where for example, information is available only up to some fixed time \( s = s_0 \) which came before time \( t \), or where the effect of the information in time interval \((s, t)\) is to be investigated. Even though this may be an interesting case, it will not be discussed in this dissertation.

The following definition of some forecasting aspects is given informally by Lütkepohl (1991: 27), and is formalized here for convenience.

**Definition (Forecasting concepts)**

1. The set \( U_t \) given by (3.2.1) is the information set.
2. The period \( t \) from where the forecast is made, is the forecast origin.
3. The number of periods into the future for which a forecast is required, is the forecast horizon, and lastly
4. A predictor \( t \) steps (periods) ahead is called an \( t \)-step ahead forecast.

We judge the optimality of forecasts by using the mean square error (MSE) matrix, which will be defined shortly. Granger (1969), Granger & Newbold (1986) and Lütkepohl show that minimum MSE forecasts minimize other functions as well, functions which may be used as criteria functions. The previous comment will be revisited after the definition.

**Definition (MSE)**


Let \( \begin{bmatrix} x_{1t} \\
 x_{2t} \end{bmatrix} \) be a time series process and \( \begin{bmatrix} x_{1t}(\ell) \\
 x_{2t}(\ell) \end{bmatrix} \) any \( \ell \)-step ahead forecast at forecast origin \( t \). The MSE predictor for \( \begin{bmatrix} x_{1t}(\ell) \\
 x_{2t}(\ell) \end{bmatrix} \) is:

\[
\text{MSE} \begin{bmatrix} x_{1t}(\ell) \\
 x_{2t}(\ell) \end{bmatrix} = E \begin{bmatrix} x_{1,t+\ell} - x_{1t}(\ell) \\
 x_{2,t+\ell} - x_{2t}(\ell) \end{bmatrix} \begin{bmatrix} x_{1,t+\ell} - x_{1t}(\ell) \\
 x_{2,t+\ell} - x_{2t}(\ell) \end{bmatrix}'
\]

(3.2.2)

It is desirable to find those forecasts \( \begin{bmatrix} x_{1t}(\ell) \\
 x_{2t}(\ell) \end{bmatrix} \) which are close to the real \( \begin{bmatrix} x_{1,t+\ell} \\
 x_{2,t+\ell} \end{bmatrix} \), and those will reduce the differences \( \begin{bmatrix} x_{1,t+\ell} - x_{1t}(\ell) \\
 x_{2,t+\ell} - x_{2t}(\ell) \end{bmatrix} \) to small values. As a result the elements of the MSE matrix are reduced. We deduce from here that desirable forecast is that which produces minimum MSE, and this will be called minimum MSE forecast, or optimal forecast.
Matrices are compared as

\[ B > A \quad (\text{or} \quad A < B) \]

if and only if

\[ B - A \text{ is positive definite} \]

**Definition (Minimum MSE)**

Lütkepohl (1991: 28)

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a time series process. Then

1. the conditional MSE predictor for forecast horizon \( t \) at origin \( t \) is the conditional expected value

\[
E_t \left[ x_{1,t+\ell} \right] = E \left[ \begin{bmatrix} x_{1,t+\ell} \\ x_{2,t+\ell} \end{bmatrix} \mid \begin{bmatrix} x_{1s} \\ x_{2s} \end{bmatrix} \mid s \leq t \right] \tag{3.2.3}
\]

and

2. the corresponding minimum MSE matrix is given by

\[
\text{MSE} \left[ E_t \left[ x_{1,t+\ell} \right] = E \left[ \begin{bmatrix} x_{1,t+\ell} - E x_{1,t+\ell} \\ x_{2,t+\ell} - E x_{2,t+\ell} \end{bmatrix} \begin{bmatrix} x_{1,t+\ell} - E x_{1,t+\ell} \\ x_{2,t+\ell} - E x_{2,t+\ell} \end{bmatrix} \right] \right] \tag{3.2.4}
\]

Forecast properties of bivariate time series are affected by different conditions such as the relation between the two components (inter alia). We have in mind causality and cointegration (as mentioned earlier). The rest of this chapter treats causality, and cointegration is left for Chapter 4.

### 3.3 GRANGER'S CONCEPT OF CAUSALITY

Harvey (1981a: 300-307)

The essence of Granger's idea is that \( z \) causes \( y \) if consideration of past (and sometimes present) values of \( z \) leads to improved predictions or forecasts for \( y \). The idea has in it, from 'past values of \( z \)', that a cause
cannot come after the effect. In addition to the information set $U_t$ of (3.2.1), for any process $z_t$ (univariate or bivariate), let the optimum $t$-step forecast at origin $t$ be denoted

$$x_t \left[ t \mid \left\{ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right\} \mid s \leq t \right]$$

and the corresponding forecast MSE by

$$\Sigma_{zz} \left[ t \mid \left\{ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right\} \mid s \leq t \right]$$

Let $[x_{1t}]$ be our time series and $\sigma_{ii}(\ell)$ be the $\ell$-step forecast MSE for $z_{it}$, $i = 1, 2$. For convenience we define "$x_{1t}$ causes $x_{2t}$" with the understanding that the reverse is defined analogously.

Definition (Granger-causality, instantaneous causality)

1. Harvey (1990: 304)
   The process $x_{1t}$ is said to cause $x_{2t}$ in Granger's sense if
   $$\sigma_{22} \left[ t \mid \left\{ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right\} \mid s \leq t \right] < \sigma_{22} \left[ t \mid \left\{ x_{2s} \mid s \leq t \right\} \right] \quad (3.3.1)$$

   The process $x_{1t}$ is said to cause $x_{2t}$ instantaneously if
   $$\sigma_{22} \left[ 1 \mid \left\{ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right\} \mid s \leq t \right] < \sigma_{22} \left[ 1 \mid \left\{ x_{2s} \mid s \leq t \right\} \right] \quad (3.3.2)$$
3. The process \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) is called a feedback system if \( x_{1t} \) causes \( x_{2t} \), and \( x_{2t} \) causes \( x_{1t} \), both in Granger's sense.

3.3.1 Characterization of Granger-causality for bivariate time series

Lütkepohl (1991: 37); Discussed with Prof. Markham

We write (3.1.10) as

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \varepsilon_{1t} + \sum_{i=1}^{\infty} \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\ \psi_{21,i} & \psi_{22,i} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix}
\]

(3.3.3a)

or

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}
\]

(3.3.3b)

where

\[
\begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_{i=1}^{\infty} \begin{bmatrix} \psi_{11,i}L^i & \psi_{12,i}L^i \\ \psi_{21,i}L^i & \psi_{22,i}L^i \end{bmatrix}
\]

Now

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \varepsilon_{1t} + \sum_{i=1}^{\infty} \psi_{11,i}\varepsilon_{1,t-i} + \sum_{i=1}^{\infty} \psi_{12,i}\varepsilon_{2,t-i}
\]

As we know that \( \psi_{11,0} = 1 \) we write

\[
x_{1t} = \sum_{i=0}^{\infty} \psi_{11,i}\varepsilon_{1,t-i} + \sum_{i=1}^{\infty} \psi_{12,i}\varepsilon_{2,t-i}
\]

(3.3.4)

The \( \ell \)-step ahead forecast of \( x_{t} \) (or the estimator of \( x_{t+\ell} \)) is
By using (3.3.3a)

\[
\begin{bmatrix}
x_{1,t+t} \\
x_{2,t+t}
\end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix}
\psi_{11,i} & \psi_{12,i} \\
\psi_{21,i} & \psi_{22,i}
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1,t+t} \\
\varepsilon_{2,t+t}
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
\psi_{11,0} & \psi_{12,0} \\
\psi_{21,0} & \psi_{22,0}
\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

Now \( \varepsilon_{t+t} \), \( t > 0 \) is independent of present and past \( x_t, x_{t-1}, \ldots \) so

\[
E(\varepsilon_{t+t} | x_t, x_{t-1}, \ldots) = 0, \ t > 0
\]

Thus we find that the minimum MSE matrix predictor of \( x_{t+t} \) based on \( x_t, x_{t-1}, \ldots \) can be represented as

\[
\begin{bmatrix}
x_{1,t+t} \\
x_{2,t+t}
\end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix}
\psi_{11,i+t} & \psi_{12,i+t} \\
\psi_{21,i+t} & \psi_{22,i+t}
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1,t-i} \\
\varepsilon_{2,t-i}
\end{bmatrix}
\]

The optimal 1-step ahead forecast of \( x_{1t} \) based on \( x_{1t}, x_{2t} \) is

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\begin{bmatrix}
[z_{1s}] | s \leq t \\
[z_{2s}]
\end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_{1t} \\
x_{2t}
\end{bmatrix}
\]

\[ x_{1t}^{(1)} \left| \begin{bmatrix} x_{1s} & x_{2s} \end{bmatrix} \right| s \leq t \] = \sum_{i=1}^{\infty} \psi_{11,i}^{(1)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(1)} \epsilon_{2,t+1-i} + \sum_{i=1}^{\infty} \psi_{11,i}^{(2)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(2)} \epsilon_{2,t+1-i}

(3.3.5)

That is

\[ x_{1t}^{(1)} \left| \begin{bmatrix} x_{1s} & x_{2s} \end{bmatrix} \right| s \leq t \] = \sum_{i=1}^{\infty} \psi_{11,i}^{(1)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(1)} \epsilon_{2,t+1-i} + \sum_{i=1}^{\infty} \psi_{11,i}^{(2)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(2)} \epsilon_{2,t+1-i}

From (3.3.4) and (3.3.5) we evaluate the forecast error

\[ x_{1,t+1} - x_{1t}^{(1)} \left| \begin{bmatrix} x_{1s} & x_{2s} \end{bmatrix} \right| s \leq t \] = \sum_{i=1}^{\infty} \psi_{11,i}^{(1)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(1)} \epsilon_{2,t+1-i} + \sum_{i=1}^{\infty} \psi_{11,i}^{(2)} \epsilon_{1,t+1-i} + \sum_{i=1}^{\infty} \psi_{12,i}^{(2)} \epsilon_{2,t+1-i}

(3.3.6)

That is

\[ x_{1,t+1} - x_{1t}^{(1)} \left| \begin{bmatrix} x_{1s} & x_{2s} \end{bmatrix} \right| s \leq t \] = \epsilon_{1,t+1}

Lütkepohl (1991: 21)

Each subprocess of a stationary process has a prediction error MA representation

\[ x_{1t}^{(1)} = \sum_{i=0}^{\infty} f_{i} v_{t-i} \]

(3.3.7)

where \( f_{0} = 1 \) and \( \{v_{t}\} \) is a white noise process.

Thus the optimal 1-step ahead predictor based on \( x_{1t} \) only, is

\[ x_{1t}^{(1)}(1 \mid \{x_{1s} \mid s \leq t\}) = x_{1t}^{(1)} \]
The corresponding 1-step ahead forecast error is
\[ x_{1,t+1} - x_{1t}(1) = v_{t+1} \] (3.3.9)

The predictors (3.3.5) and (3.3.8) are identical if and only if
\[ v_t = \varepsilon_{1t} \] for all \( t \),
that is, equality of predictors is equivalent to \( x_{1t} \) having the prediction error MA representation
\[
x_{1t} = \sum_{i=1}^{\infty} f_i v_{t+1-i}
\]
\[ = \sum_{i=0}^{\infty} \begin{bmatrix} f_i & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-i} \\ \varepsilon_{2,t-i} \end{bmatrix} = \sum_{i=0}^{\infty} \psi_{11,i} \varepsilon_{1,t-i} + \sum_{i=1}^{\infty} \psi_{12,i} \varepsilon_{2,t-i} \quad : \psi_{12,0} = 0
\]

Since the prediction error MA representation is unique, then the above equality is possible if and only if
\[ f_i = \psi_{11,i} \]
\[ \psi_{12,i} = 0 \]
for all \( i = 1, 2, 3, ... \)

We have proved the following theorem.
Theorem 3.2

Lütkepohl (1991: 38 Proposition 2.2)

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate process as in (3.3.3) with prediction error operator

\[
\begin{bmatrix}
\psi_{11}(z) & \psi_{12}(z) \\
\psi_{21}(z) & \psi_{22}(z)
\end{bmatrix}
\]

Then

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
z_{1s} \\
z_{2s}
\end{bmatrix}
\end{bmatrix} = z_{1t}(1 \{ z_{1s} \leq t \})
\]

if and only if

\[
\psi_{12,i} = 0 \text{ for } i = 1, 2, ...
\]

The following corollaries will be proved based on the above theorem.

Corollary 3.2.1

Lütkepohl (1991: 39 Corollary 2.2.1)

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate AR process as in (3.3.3) with prediction error MA operator \( \begin{bmatrix}
\psi_{11}(z) & \psi_{12}(z) \\
\psi_{21}(z) & \psi_{22}(z)
\end{bmatrix} \). Then

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
z_{1s} \\
z_{2s}
\end{bmatrix}
\end{bmatrix} = z_{1t}(1 \{ z_{1s} \leq t \})
\]

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
z_{1s} \\
z_{2s}
\end{bmatrix}
\end{bmatrix} = z_{1t}(1 \{ z_{1s} \leq t \})
\]

if and only if

\[
\psi_{12,i} = 0 \text{ for } i = 1, 2, ...
\]

Proof

Let the information set be \( \begin{bmatrix} z_{1s} \leq t \\ z_{2s} \end{bmatrix} \), then using the origin \( t \), the \( \ell \)-step ahead forecast \( \begin{bmatrix} x_{1t}(\ell) \\ x_{2t}(\ell) \end{bmatrix} \) is a 1-step ahead forecast for \( \begin{bmatrix} x_{1t}(\ell-1) \\ x_{2t}(\ell-1) \end{bmatrix} \).
where \( t \geq 1 \), and 
\[
\begin{bmatrix}
    x_{1t}(0) \\
    x_{2t}(0)
\end{bmatrix} =
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
\]

Using (3.3.3) and the above statements, equality of 1-step ahead predictors implies equality of the \( \ell \)-step forecasts for \( \ell = 2, 3, \ldots \). Hence, from (3.3.10) we have that

\[
x_{1t} \left( \ell \right| \begin{bmatrix} x_{1s} \mid s \leq t \end{bmatrix} \right) = x_{1t} \left( \ell \right| \{ x_{1s} \mid s \leq t \})
\]

if and only if

\[
\psi_{12,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots
\]

That is

\[
x_{2t} \quad \text{does not lead to improved forecasts of} \quad x_{1t}
\]

if and only if

\[
\psi_{12,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots
\]

Therefore,

\[
x_{2t} \quad \text{does not Granger-cause} \quad x_{1t}
\]

if and only if

\[
\psi_{12,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots
\]

\[\square\]

**Corollary 3.2.2**

Lütkepohl (1991: 39 Corollary 2.2.1)

Suppose 
\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
\]

is a stable bivariate AR(p) process

\[
\begin{bmatrix}
    z_{1t} \\
    z_{2t}
\end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix}
    \phi_{11,i} & \phi_{12,i} \\
    \phi_{21,i} & \phi_{22,i}
\end{bmatrix} \begin{bmatrix}
    x_{1,t-i} \\
    x_{2,t-i}
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

with white noise \( \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix} \) of nonsingular covariance matrix \( \Sigma_{\varepsilon \varepsilon} \). Then

\[
\begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]
\[ x_{1t} [ \ell | \{ x_{1s} \mid s \leq t \} ] = x_{1t} ( \ell | \{ x_{1s} \mid s \leq t \} ), \quad \ell = 1, 2, \ldots \]  \tag{3.3.12a}

if and only if
\[ \phi_{12,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots, p \]  \tag{3.3.12b}

**Proof**

(This proof was discussed with Prof. Markham)

Writing (3.3.11) as

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

with

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\phi_{11,i} & \phi_{12,i} \\
\phi_{21,i} & \phi_{22,i}
\end{bmatrix}
\]

Under the given conditions we have shown that using SSR that \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) has the prediction error MA representation

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

where now (3.3.12) is valid if and only if \( \psi_{12}(L) = 0 \) from Corollary 3.2.1.

Lütkepohl (1991: 17) states that when \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) is stable, then \( \Phi(L) \) is nonsingular and is the inverse of \( \psi(L) \), that is (3.3.12) is valid under the condition

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\psi_{11}(L) & 0 \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}^{-1}
\]
That is, $x_{1t}$ is not Granger-caused by $x_{2t}$ if and only if
\[ \phi_{12}(L) = 0 \]
But
\[ \phi_{12}(L) = \sum_{i=1}^{p} \phi_{12,i} L^i \]
and therefore
\[ \phi_{12}(L) = 0 \text{ implies } \phi_{12,1} = \ldots = \phi_{12,p} = 0 \]
Therefore, (3.3.12) is valid.

3.3.2 Characterization of instantaneous causality
Let us consider the prediction error MA representation (3.3.3) which is
\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= \begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix}
+ \sum_{i=1}^{\infty} \begin{bmatrix}
    \psi_{11,i} & \psi_{12,i} \\
    \psi_{21,i} & \psi_{22,i}
\end{bmatrix}
\begin{bmatrix}
    \epsilon_{1,t-1} \\
    \epsilon_{2,t-1}
\end{bmatrix}
\]
where
\[
\sum_{\epsilon\epsilon} = \begin{bmatrix}
    \sigma_{11,\epsilon} & \sigma_{12,\epsilon} \\
    \sigma_{21,\epsilon} & \sigma_{22,\epsilon}
\end{bmatrix}
\]
is positive definite.

Harvey (1989: 131) and Lütkepohl (1991: 462) conclude that every positive definite matrix has a Cholesky decomposition, this means that there is a matrix $L,$
\[
L = \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}, \quad \ell_{11} > 0, \quad \ell_{22} > 0
\]
such that
\[
\begin{bmatrix}
    \sigma_{11,\epsilon} & \sigma_{12,\epsilon} \\
    \sigma_{21,\epsilon} & \sigma_{22,\epsilon}
\end{bmatrix}
= \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}
\begin{bmatrix}
    \ell_{11} & \ell_{21} \\
    0 & \ell_{22}
\end{bmatrix}
\]
Lütkepohl (1991: 40 (2.3.15))
We write (with known notation)

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix}
    \psi_{11,i} & \psi_{12,i} \\
    \psi_{21,i} & \psi_{22,i}
\end{bmatrix} \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix} \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}^{-1} \begin{bmatrix}
    \varepsilon_{1,t-i} \\
    \varepsilon_{2,t-i}
\end{bmatrix}
\]

and define

\[
\begin{bmatrix}
    \theta_{11,i} & \theta_{12,i} \\
    \theta_{21,i} & \theta_{22,i}
\end{bmatrix} = \begin{bmatrix}
    \psi_{11,i} & \psi_{12,i} \\
    \psi_{21,i} & \psi_{22,i}
\end{bmatrix} \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    \psi_{11,i}\ell_{11} + \psi_{12,i}\ell_{21} & \psi_{12,i}\ell_{22} \\
    \psi_{21,i}\ell_{11} & \psi_{22,i}\ell_{21} + \psi_{22,i}\ell_{22}
\end{bmatrix}
\]

for \( i = 0, 1, 2, \ldots \)

Since

\[
\begin{bmatrix}
    \psi_{11,0} & \psi_{12,0} \\
    \psi_{21,0} & \psi_{22,0}
\end{bmatrix} = \begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix}
\]

then

\[
\begin{bmatrix}
    \theta_{11,0} & \theta_{12,0} \\
    \theta_{21,0} & \theta_{22,0}
\end{bmatrix} = \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}
\]

so that

\[
\theta_{12,0} = 0
\]

Define

\[
\begin{bmatrix}
    w_{1t} \\
    w_{2t}
\end{bmatrix} = \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}^{-1} \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

then for each integer \( t \),

\[
E \begin{bmatrix}
    w_{1t} \\
    w_{2t}
\end{bmatrix} = \begin{bmatrix}
    0 \\
    0
\end{bmatrix}
\]

and

\[
E \begin{bmatrix}
    w_{1t} \quad w_{1t} & w_{2t} \\
    w_{2t}
\end{bmatrix} = \begin{bmatrix}
    \ell_{11} & 0 \\
    \ell_{21} & \ell_{22}
\end{bmatrix}^{-1} \begin{bmatrix}
    \sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\
    \sigma_{21,\varepsilon} & \sigma_{22,\varepsilon}
\end{bmatrix} \begin{bmatrix}
    \ell_{11} & \ell_{21} \\
    \ell_{21} & \ell_{22}
\end{bmatrix}^{-1}
\]
Definition (Orthogonal residuals/innovations)

Lütkepohl (1991: 40)

The white noise errors \( \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} \) with uncorrelated components defined in (3.3.15) are known as orthogonal residuals or innovations.

Lütkepohl (1991: 40)

The MA representation with orthogonal innovations is therefore

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \theta_{11,0} & 0 \\ \theta_{21,0} & \theta_{22,0} \end{bmatrix} \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11,1} & \theta_{12,1} \\ \theta_{21,1} & \theta_{22,1} \end{bmatrix} \begin{bmatrix} w_{1,t-1} \\ w_{2,t-1} \end{bmatrix} + \ldots
\]

(3.3.16)

which implies that

\[
x_{1,t+1} = \theta_{11,0} w_{1,t+1} + \sum_{i=1}^{\infty} \theta_{11,i} w_{1,t+1-i} + \sum_{i=1}^{\infty} \theta_{12,i} w_{2,t+1-i}
\]

(3.3.17a)

and

\[
x_{2,t+1} = \theta_{21,0} w_{2,t+1} + \theta_{22,0} w_{2,t+1} + \sum_{i=1}^{\infty} \theta_{21,i} w_{1,t+1-i} + \sum_{i=1}^{\infty} \theta_{22,i} w_{2,t+1-i}
\]

(3.3.17b)

The following discussion was done with the help from Prof. Markham. Outline is given in Lütkepohl (1991: 40).

From (3.3.17) we note that the processes \((x_{1t})\) and \((x_{2t})\) are functions of the innovations \((w_{1t})\) and \((w_{2t})\). Also, at time \(t\) we have each of \(x_{1t}\) and \(x_{2t}\) being given by \(w_{1t}, w_{2t}, w_{1,t-1}, w_{2,t-1}, \ldots\); \(x_{1,t+1}\) and \(x_{2,t+1}\) are
based on \( w_{1,t+1}, w_{2,t+1}, w_{1t}, w_{2t}, \ldots \) and so on. That is, the information about \( x_{1,t+1} \) or \( x_{2,t+1} \) can be fully given using up to \( w_{1,t+1} \) and \( w_{2,t+1} \). Then, based on the information set \( \{ x_{1s} \mid s \leq t \} \) and on \( x_{1,t+1} \), the 1-step ahead predictor of \( x_{2t} \) is equal to 1-step ahead predictor of \( x_{2t} \) based on the information set \( \{ x_{1s} \mid s \leq t \} \) and on \( w_{1,t+1} \). That is,

\[
\hat{x}_{2t} \left( 1 \mid \left[ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right] \cup \{ x_{1,t+1} \} \right) = \hat{x}_{2t} \left( 1 \mid \left[ \begin{array}{c} w_{1s} \\ w_{2s} \end{array} \right] \cup \{ w_{1,t+1} \} \right)
\]

(3.3.18)

For \( s \leq t \),

\[
E \left( \begin{array}{c} w_{1,t+1} \\ w_{1s} \\ w_{2s} \end{array} \right) = \left[ \begin{array}{c} 0 \\ 0 \end{array} \right]
\]

so that

\[
x_{2t} \{ 1 \mid \{ w_{1,t+1} \} \} = E(x_{2,t+1} \mid \{ w_{1,t+1} \}) = \theta_{21,0} w_{1,t+1}
\]

The right hand side of (3.3.18) is

\[
x_{2t} \left( 1 \mid \left[ \begin{array}{c} w_{1s} \\ w_{2s} \end{array} \right] \cup \{ w_{1,t+1} \} \right) = x_{2t} \{ 1 \mid \{ w_{1,t-1} \} \} + x_{2t} \left( 1 \mid \left[ \begin{array}{c} w_{1s} \\ w_{2s} \end{array} \right] \cup \{ w_{1,t+1} \} \right)
\]

\[
= \theta_{21,0} w_{1,t+1} + x_{2t} \left( 1 \mid \left[ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right] \right)
\]

The equation (3.3.18) becomes

\[
x_{2t} \left( 1 \mid \left[ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right] \cup \{ x_{1,t+1} \} \right) = \theta_{21,0} w_{1,t+1} + x_{2t} \left( 1 \mid \left[ \begin{array}{c} x_{1s} \\ x_{2s} \end{array} \right] \right)
\]
From (3.3.14) we conclude that (3.3.19) is true if and only if

\[ L = \begin{bmatrix} \ell_{11} & 0 \\ 0 & \ell_{22} \end{bmatrix} \]

and this is when

\[
\begin{bmatrix}
\sigma_{11,e} & \sigma_{12,e} \\
\sigma_{21,e} & \sigma_{22,e}
\end{bmatrix} = \begin{bmatrix} \ell_{11}^2 & 0 \\ 0 & \ell_{22}^2 \end{bmatrix}
\]

That is, \( \sigma_{12,e} = \mathbb{E}(\varepsilon_{1t}\varepsilon_{2t}) = 0 \)

That is,

\[
x_{2t} \begin{bmatrix} 1 \\ \{ [x_{1s} | s \leq t] \} \cup \{ x_{1,t+1} \} \end{bmatrix} = x_{2t} \begin{bmatrix} 1 \\ \{ [x_{1s} | s \leq t] \} \cup \{ x_{2s} \} \end{bmatrix}
\]

if and only if

\[ \mathbb{E}(\varepsilon_{1t}\varepsilon_{2t}) = 0 \]

We have proved the following theorem.

**Theorem 3.3 (Characterization of instantaneous causality)**

Lütkepohl (1991: 41 Proposition 2.3)

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be as in (3.3.16) with nonsingular innovation covariance matrix
\[\Sigma_{ee} \] Then there is no instantaneous causality between \(x_{1t}\) and \(x_{2t}\) if and only if \(\text{E}(\varepsilon_{1t}, \varepsilon_{2t}) = 0\).

Theorems 3.2 and 3.3 are both proved in terms of the prediction error MA representation, but the former started from a bivariate AR(p). Also, the latter is in terms of the elements of the MA operator, but of course extended in terms of the AR operator in Corollary 3.2.2, whereas the former is in terms of the error terms. The following corollary of the above theorem shows the instantaneous causality on a bivariate time series, that is, 'what happens if there is no instantaneous causality between \(x_{1t}\) and \(x_{2t}\).

**Corollary**

Lütkepohl (1991: 41)

Let \(\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}\) be a stable bivariate AR(p) process as in (3.3.11) with a nonsingular white noise covariance matrix \(\Sigma_{ee} = \begin{bmatrix} \sigma_{11,e} & \sigma_{12,e} \\ \sigma_{21,e} & \sigma_{22,e} \end{bmatrix}\). Then there is no instantaneous causality between \(x_{1t}\) and \(x_{2t}\) if and only if

\[\sigma_{12,e} = \text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = 0.\]

The following proof was suggested by Prof. Markham. Also, Lütkepohl (1991: 37-41).

**Proof of Corollary**

The conditions we have are as in (3.1.5) and from (3.1.5) to (3.1.9) we show this stability AR(p) has the MA representation

\[
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} = \begin{bmatrix}
\psi_{11,i} & \psi_{12,i} \\
\psi_{21,i} & \psi_{22,i}
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1,t-i} \\
\varepsilon_{2,t-i}
\end{bmatrix} + \sum_{i=1}^{\infty} \begin{bmatrix}
\sigma_{11,e} & \sigma_{12,e} \\
\sigma_{21,e} & \sigma_{22,e}
\end{bmatrix}
\]

Using the Cholesky decomposition (Harvey 1989: 131, Lütkepohl 1991: 462)
\[
\begin{bmatrix}
\sigma_{11,e} & \sigma_{12,e} \\
\sigma_{21,e} & \sigma_{22,e}
\end{bmatrix} =
\begin{bmatrix}
\ell_{11} & 0 \\
\ell_{21} & \ell_{22}
\end{bmatrix}
\begin{bmatrix}
\ell_{11} \\
\ell_{21} \\
0 \\
\ell_{22}
\end{bmatrix},
\]

also (3.3.13) is
\[
\begin{bmatrix}
\theta_{11,i} & \theta_{12,i} \\
\theta_{21,i} & \theta_{22,i}
\end{bmatrix} =
\begin{bmatrix}
\psi_{11,i} & \psi_{12,i} \\
\psi_{21,i} & \psi_{22,i}
\end{bmatrix}
\begin{bmatrix}
\ell_{11} & 0 \\
\ell_{21} & \ell_{22}
\end{bmatrix}
\text{for } i = 0, 1, 2, ...
\]

and (3.3.15) is
\[
\begin{bmatrix}
w_{1t} \\
w_{2t}
\end{bmatrix} =
\begin{bmatrix}
\ell_{11} & 0 \\
\ell_{21} & \ell_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\text{for all } t.
\]

We have represented the above MA in (3.3.16) as
\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
\theta_{11,0} & \theta_{12,0} \\
\theta_{21,0} & \theta_{22,0}
\end{bmatrix}
\begin{bmatrix}
w_{1t} \\
w_{2t}
\end{bmatrix}
+ \begin{bmatrix}
\theta_{11,1} & \theta_{12,1} \\
\theta_{21,1} & \theta_{22,1}
\end{bmatrix}
\begin{bmatrix}
w_{1,t-1} \\
w_{2,t-1}
\end{bmatrix} + \ldots
\]

Using Theorem 3.3 there is therefore no instantaneous causality if and only if
\[
\sigma_{12,t} = \text{cov}(\epsilon_{1t}, \epsilon_{2t}) = 0
\]

3.4 GRANGER-CAUSALITY OF STABLE ARMA(p,q)
Lütkepohl (1991: 17)

Definition (Invertibility)
Let the ARIMA(p,d,q) (3.1.1) be given. The AR operator
\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\]

is said to be invertible if
The bivariate model \[
\begin{bmatrix}
\phi_{11}(z) & \phi_{12}(z) \\
\phi_{21}(z) & \phi_{22}(z)
\end{bmatrix}
\neq 0 \text{ for } |z| \leq 1.
\]
The bivariate model \[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\]
with invertible AR operator is called invertible process.

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}^{-1} = \begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}
\]
exists where

\[
\psi_{ij}(L) = \sum_{k=0}^{\infty} \psi_{ij,k} L^k, \quad i, j = 1, 2
\]

and

\[
\sum_{k=0}^{\infty} |\psi_{ij,k}| < \infty.
\]

The prediction error MA of stable processes is very important, and through them we derived conditions of causality of AR processes. We proceed along this approach on the ARMA processes, starting with a simpler ARMA(1,1) process.

### 3.4.1 ARMA(1,1) process as infinite order MA process


Suppose that the bivariate ARMA(1,1) process

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\phi_{11} & 0 \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} + \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{21} & \theta_{22}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-1} \\
\varepsilon_{2,t-1}
\end{bmatrix}
\]

is stable, that is

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\phi_{11} & 0 \\
\phi_{21} & \phi_{22}
\end{bmatrix} z \neq 0 \text{ if } |z| \leq 1
\]
then

\[ \Phi(z) = \begin{bmatrix} \phi_{11}(z) & \phi_{12}(z) \\ \phi_{21}(z) & \phi_{22}(z) \end{bmatrix} \]

is invertible, where

\[ \phi_{11}(z) = 1 - \phi_{11}z \quad \phi_{12}(z) = 0 \]
\[ \phi_{21}(z) = -\phi_{21}z \quad \phi_{22}(z) = 1 - \phi_{22}z \]

and the inverse is

\[
\begin{bmatrix}
1 - \phi_{11}z & 0 \\
-\phi_{21}z & 1 - \phi_{22}z
\end{bmatrix}^{-1} = \frac{1}{(1-\phi_{11}z)(1-\phi_{22}z)} \begin{bmatrix}
1 - \phi_{22}z & 0 \\
\phi_{21}z & 1 - \phi_{11}z
\end{bmatrix}
\]

where

\[
\frac{\phi_{21}z}{(1-\phi_{11}z)(1-\phi_{22}z)} = \frac{\phi_{21}/(\phi_{11} - \phi_{22})}{1-\phi_{11}z} - \frac{\phi_{21}/(\phi_{11} - \phi_{22})}{1-\phi_{22}z}
\]

Since \(|\phi_{11}|, |\phi_{22}| < 1\) as absolute values of eigenvalues of a stable process, we choose \(z\) such that \(|z| \leq 1\), then

\[
\frac{1}{1-\phi_{11}z} = 1 + \phi_{11}z + \phi_{11}^2z^2 + \ldots
\]
\[
\frac{1}{1-\phi_{22}z} = 1 + \phi_{22}z + \phi_{22}^2z^2 + \ldots
\]

We define

\[ \pi_{ii,k} = \phi_{ii}^k, \quad i = 1, 2, \text{ and } k = 0, 1, 2, \ldots \]
\[ \pi_{21,0} = 0, \quad \pi_{21,1} = \phi_{21}, \quad \pi_{21,2} = \phi_{21}(\phi_{11} + \phi_{22}), \quad \text{and so on} \]

and

\[ \pi_{12,k} = 0 \quad \text{for all} \quad k \]

and now we write

\[
\begin{pmatrix}
1 - \phi_{11}z & 0 \\
- \phi_{21}z & 1 - \phi_{22}z
\end{pmatrix}^{-1} =
\begin{pmatrix}
\pi_{11}(z) & \pi_{12}(z) \\
\pi_{21}(z) & \pi_{22}(z)
\end{pmatrix} \tag{3.3.21}
\]

where

\[ \pi_{12}(z) = 0 \]

\[ \pi_{ii}(z) = \sum_{k=0}^{\infty} \pi_{ii,k}z^k, \quad i = 1, 2 \]

and

\[ \pi_{21}(z) = \sum_{k=1}^{\infty} \pi_{21,k}z^k, \quad \text{since} \quad \pi_{21,0} = 0 \]

Let us define

\[ \Theta(L) =
\begin{pmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{pmatrix}
\]

\[ =
\begin{pmatrix}
1 - \theta_{11}L & - \theta_{12}L \\
- \theta_{21}L & 1 - \theta_{22}L
\end{pmatrix}
\]

then we write (3.3.20) as

\[
\begin{pmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{pmatrix}
\begin{pmatrix}
x_{1t} \\
x_{2t}
\end{pmatrix} =
\begin{pmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{pmatrix}
\]

Premultiplying both sides by (3.3.21) gives

\[
\begin{pmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{pmatrix}^{-1}
\begin{pmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{pmatrix}
\begin{pmatrix}
x_{1t} \\
x_{2t}
\end{pmatrix}
\]
\[
\begin{bmatrix}
\pi_{11}(L) & \pi_{12}(L) \\
\pi_{21}(L) & \pi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

so that

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\pi_{11}(L)\theta_{11}(L)+\pi_{12}(L)\theta_{21}(L) & \pi_{11}(L)\theta_{12}(L)+\pi_{12}(L)\theta_{22}(L) \\
\pi_{21}(L)\theta_{11}(L)+\pi_{22}(L)\theta_{21}(L) & \pi_{21}(L)\theta_{12}(L)+\pi_{22}(L)\theta_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

where

\[
\pi_{11}(L)\theta_{11}(L)+\pi_{12}(L)\theta_{12}(L) = \left[ \sum_{i=0}^{\infty} \pi_{11,i} L^i \right] (1 - \theta_{11} L) : \pi_{12}(L) = 0
\]

\[
= \sum_{i=0}^{\infty} \psi_{11,i} L^i
\]

with

\[
\psi_{11,0} = \pi_{11,0} \quad \text{and} \quad \psi_{11,i} = \pi_{11,i} - \theta_{11} \pi_{11,i-1} \quad \text{for } i = 1, 2, ...
\]

Similarly

\[
\pi_{21}(L)\theta_{11}(L)+\pi_{22}(L)\theta_{21}(L) = \left[ \sum_{i=0}^{\infty} \pi_{21,i} L^i \right] (1 - \theta_{11} L) + \left[ \sum_{i=0}^{\infty} \pi_{21,i} L^i \right] (-\theta_{21} L)
\]

\[
= \pi_{21,0} + \sum_{i=1}^{\infty} \left( \pi_{21,i} - \theta_{22,i-1} - \theta_{21} \pi_{22,i-1} \right) L^i
\]

\[
= \sum_{i=0}^{\infty} \psi_{21,i} L^i
\]

with

\[
\psi_{21,0} = \pi_{21,0} \quad \text{and} \quad \psi_{21,i} = \pi_{21,i} - \theta_{22} \pi_{21,i-1} - \theta_{21} \pi_{22,i-1}
\]

\[
\text{for } i = 1, 2, ...
\]

Next

\[
\pi_{11}(L)\theta_{12}(L) + \pi_{12}(L)\theta_{22}(L) = \left[ \sum_{i=0}^{\infty} \pi_{11,i} L^i \right] (-\theta_{12} L) : \pi_{12}(L) = 0
\]

\[
= \sum_{i=0}^{\infty} \left( -\pi_{11,i} \theta_{12} \right) L^{i+1}
\]
with $\psi_{12,0} = 0$ and $\psi_{12,i} = -\pi_{11,i-1} \theta_{12}$ for $1 = 1, 2, ...$

and lastly

$$\pi_{21}(L)\theta_{12}(L) + \pi_{22}(L)\theta_{22}(L) = \left[\sum_{i=0}^{\infty} \pi_{21,i} L^i\right](-\theta_{12} L) + \left[\sum_{i=0}^{\infty} \pi_{22,i} L^i\right](1-\theta_{22} L)$$

$$: \pi_{21,0} = 0$$

$$= \pi_{22,0} + (\pi_{21,1} - \pi_{22,0} \theta_{22}) L + \sum_{i=2}^{\infty} (\pi_{22,i} - \pi_{22,i-1} \theta_{22} - \pi_{21,i-1} \theta_{12}) L^i$$

$$= \sum_{i=0}^{\infty} \psi_{22,i} L^i$$

with

$$\psi_{22,0} = \pi_{22,0}, \quad \psi_{22,1} = \pi_{21,1} - \pi_{22,0} \theta_{22}$$

$$\psi_{22,i} = \pi_{22,i} - \pi_{22,i-1} \theta_{22} - \pi_{21,i-1} \theta_{12}, \quad i = 1, 2, ...$$

By defining

$$\psi_{pq}(L) = \sum_{i=0}^{\infty} \psi_{pq,i} L^i, \quad p, q = 1, 2$$

we can write (3.3.20) as

$$\begin{bmatrix} \sigma_1 \chi_1 \n \sigma_2 \chi_2 \end{bmatrix} = \begin{bmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

$$= \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11,i} L^i & \psi_{12,i} L^i \\ \psi_{21,i} L^i & \psi_{22,i} L^i \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

$$= \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{11,i} & \psi_{12,i} \\ \psi_{21,i} & \psi_{22,i} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-i} \\ \epsilon_{2,t-i} \end{bmatrix}$$

(3.3.22)
Lütkepohl (1991: 218-9)
If we assume that (3.3.20) is an invertible process, then (3.3.22) is known as the prediction error MA representation of (3.3.20).

3.4.2 Bivariate ARMA(p,q) as prediction error MA representation
The bivariate ARMA(p,q) process is

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    \phi_{11,1} & \phi_{12,1} \\
    \phi_{21,1} & \phi_{22,1}
\end{bmatrix} \begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix} + \cdots + \begin{bmatrix}
    \phi_{11,p} & \phi_{12,p} \\
    \phi_{21,p} & \phi_{22,p}
\end{bmatrix} \begin{bmatrix}
    x_{1,t-p} \\
    x_{2,t-p}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix} + \begin{bmatrix}
    \theta_{11,1} & \theta_{12,1} \\
    \theta_{21,1} & \theta_{22,1}
\end{bmatrix} \begin{bmatrix}
    \varepsilon_{1,t-1} \\
    \varepsilon_{2,t-1}
\end{bmatrix} + \cdots + \begin{bmatrix}
    \theta_{11,q} & \theta_{12,q} \\
    \theta_{21,q} & \theta_{22,q}
\end{bmatrix} \begin{bmatrix}
    \varepsilon_{1,t-q} \\
    \varepsilon_{2,t-q}
\end{bmatrix}
\]

(3.3.23)

Suppose that \( \begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} \) is a stable process, that is

\[
\begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix} - \begin{bmatrix}
    \phi_{11,1} & \phi_{12,1} \\
    \phi_{21,1} & \phi_{22,1}
\end{bmatrix} z - \cdots - \begin{bmatrix}
    \phi_{11,p} & \phi_{12,p} \\
    \phi_{21,p} & \phi_{22,p}
\end{bmatrix} z^p \neq 0 \text{ if } |z| < 1
\]

and the operator

\[
\Phi(L) = \begin{bmatrix}
    \phi_{11}(L) & \phi_{12}(L) \\
    \phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\]

is invertible, where

\[
\phi_{11}(L) = 1 - \phi_{11,1} L - \cdots - \phi_{11,p} L^p
\]
\[
\phi_{12}(L) = - \phi_{12,1} L - \cdots - \phi_{12,p} L^p
\]
\[
\phi_{21}(L) = - \phi_{21,1} L - \cdots - \phi_{21,p} L^p
\]
\[
\phi_{22}(L) = 1 - \phi_{22,1} L - \cdots - \phi_{22,p} L^p
\]

and
where
\[ \text{det} = \phi_{11}(L)\phi_{22}(L) - \phi_{12}(L)\phi_{21}(L) \]

We have seen in ARMA(1,1) that (3.3.24) can be written
\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}^{-1} = \frac{1}{\text{det}} \begin{bmatrix}
\phi_{22}(L) & -\phi_{12}(L) \\
-\phi_{21}(L) & \phi_{11}(L)
\end{bmatrix}
\]

We rewrite
\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}^{-1} = \sum_{i=0}^{\infty} \begin{bmatrix}
\pi_{11,i}L^i & \pi_{12,i}L^i \\
\pi_{21,i}L^i & \pi_{22,i}L^i
\end{bmatrix}
\]

and premultiplying this by (3.3.25) we obtain
\[
\begin{bmatrix}
\pi_{11}(L) & \pi_{12}(L) \\
\pi_{21}(L) & \pi_{22}(L)
\end{bmatrix}
\]

We rewrite
\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix} \begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{bmatrix} \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

and premultiplying this by (3.3.25) we obtain
\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\pi_{11}(L)\theta_{11}(L)+\pi_{12}(L)\theta_{21}(L) & \pi_{11}(L)\theta_{12}(L)+\pi_{12}(L)\theta_{22}(L) \\
\pi_{21}(L)\theta_{11}(L)+\pi_{22}(L)\theta_{21}(L) & \pi_{21}(L)\theta_{12}(L)+\pi_{22}(L)\theta_{22}(L)
\end{bmatrix} \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

Now,
\[
\pi_{11}(L)\theta_{11}(L) + \pi_{12}(L)\theta_{21}(L) = \left[ \sum_{i=0}^{\infty} \pi_{11,i}L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{11,j}L^j \right]
\]

\[
+ \left[ \sum_{i=0}^{\infty} \pi_{12,i}L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{21,j}L^j \right]
\]

\[
= \sum_{i=0}^{\infty} \psi_{11,i}L^i
\]
where \( \psi_{11,0} = \pi_{11,0} \theta_{11,0} + \pi_{12,0} \theta_{21,0} \) and so on.

Similarly

\[
\pi_{21}(L) \theta_{11}(L) + \pi_{22}(L) \theta_{21}(L) = \left[ \sum_{i=0}^{\infty} \pi_{21,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{11,j} L^j \right]
+ \left[ \sum_{i=0}^{\infty} \pi_{22,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{21,j} L^j \right]
= \sum_{i=0}^{\infty} \psi_{21,i} L^i
\]

\[
\pi_{11}(L) \theta_{12}(L) + \pi_{12}(L) \theta_{22}(L) = \left[ \sum_{i=0}^{\infty} \pi_{11,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{12,j} L^j \right]
+ \left[ \sum_{i=0}^{\infty} \pi_{12,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{22,j} L^j \right]
= \sum_{i=0}^{\infty} \psi_{12,i} L^i
\]

and lastly

\[
\pi_{21}(L) \theta_{12}(L) + \pi_{22}(L) \theta_{22}(L) = \left[ \sum_{i=0}^{\infty} \pi_{12,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{12,j} L^j \right]
+ \left[ \sum_{i=0}^{\infty} \pi_{22,i} L^i \right] \left[ \sum_{j=0}^{\infty} \theta_{22,j} L^j \right]
= \sum_{i=0}^{\infty} \psi_{22,i} L^i
\]

We write

\[
\begin{bmatrix}
\tau_{1t} \\
\tau_{2t}
\end{bmatrix} =
\begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]
As in ARMA(1,1) we have that if (3.3.23) is an invertible process, then (3.3.26) is known as the prediction error MA representation.

3.4.3 Granger-causality of bivariate ARMA(p,q)
Lütkepohl (1991: 236–7)
We consider the stable, invertible bivariate ARMA(p,q) process

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{bmatrix}
\begin{bmatrix}
e_{1t} \\
e_{2t}
\end{bmatrix}
\]  

(3.3.27a)

which, as shown in the previous subsection, has the prediction error MA representation

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}
\begin{bmatrix}
e_{1t} \\
e_{2t}
\end{bmatrix}
\]  

(3.3.27b)

We know from Theorem 3.2 that \( x_{1t} \) is not Granger-caused by \( x_{2t} \) if and only if

\[ \psi_{12}(L) = 0 \]

We express this condition in terms of the elements of the AR and MA operators. We have

\[
\begin{bmatrix}
\psi_{11}(L) & \psi_{12}(L) \\
\psi_{21}(L) & \psi_{22}(L)
\end{bmatrix}
= \frac{1}{\det}
\begin{bmatrix}
\phi_{22}(L) & -\phi_{12}(L) \\
-\phi_{21}(L) & \phi_{11}(L)
\end{bmatrix}
\begin{bmatrix}
\theta_{11}(L) & \theta_{12}(L) \\
\theta_{21}(L) & \theta_{22}(L)
\end{bmatrix}
\]
\[
\begin{pmatrix}
\phi_{11}(L) & \phi_{12}(L) & \phi_{13}(L) & \phi_{14}(L) \\
\phi_{21}(L) & \phi_{22}(L) & \phi_{23}(L) & \phi_{24}(L)
\end{pmatrix}
\begin{pmatrix}
x_{1t-1} \\
x_{2t-1}
\end{pmatrix}
= \begin{pmatrix}
x_{1t} \\
x_{2t}
\end{pmatrix}
\begin{pmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{pmatrix}
\]

where
\[
det = \frac{\psi_{12}(L) - \psi_{12}(L)\phi_{22}(L)}{\psi_{22}(L) - \psi_{22}(L)\phi_{22}(L)}
\]

Therefore
\[
\psi_{12}(L) = 0
\]
is equivalent to
\[
\phi_{11}(L)\phi_{22}(L) - \phi_{12}(L)\phi_{22}(L) = 0
\]
or
\[
\theta_{12}(L) = \phi_{12}(L)\phi_{22}(L)
\]

We have proved the following theorem (cf Proposition 6.3 Lütkepohl (1991: 237))

**Theorem 3.4 (Characterization of Granger-causality of ARMA(p,q)**

Let \[ x_{1t} \text{ be a stable and invertible bivariate ARMA(p,q) as in (3.3.27).} \]

Then
\[
x_{2t} \text{ does not Granger-cause } x_{1t}
\]

if and only if
\[
\theta_{12}(L) = \phi_{12}(L)\phi_{22}(L)
\]

The following example is an expansion of Remark 1 (Lütkepohl 1991: 237-8).

**Example 3.1**

Let us consider the stable and invertible bivariate ARMA(1,1) process
\[
\begin{pmatrix}
x_{1t} \\
x_{2t}
\end{pmatrix}
= \begin{pmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{pmatrix}
\begin{pmatrix}
x_{1t-1} \\
x_{2t-1}
\end{pmatrix}
+ \begin{pmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{pmatrix}
\]

which we now write as
The condition for $x_{2t}$ not Granger-causal for $x_{1t}$, (3.3.28), is in this case given by

$$\theta_{12}L = -\phi_{12}L(1 - \phi_{22}L)^{-1}(1 + \theta_{22}L)$$

which becomes

$$\theta_{12}L - \theta_{12}\phi_{22}L^2 = -\phi_{12}L - \phi_{12}\theta_{22}L^2$$

We compare by equating coefficients of corresponding powers of $L$ so that

$$\theta_{12} = -\phi_{12} \quad \text{and} \quad \theta_{12}\phi_{22} = \phi_{12}\theta_{22}$$

Therefore,

$$\theta_{12} = -\phi_{12} \quad \text{and} \quad \phi_{22} = -\theta_{22}$$

The above ARMA(1,1) process becomes

$$\begin{bmatrix} 1 - \phi_{11}L & -\phi_{12}L \\ -\phi_{21}L & 1 - \phi_{22}L \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 + \theta_{11}L & \theta_{12}L \\ \theta_{21}L & 1 + \theta_{22}L \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Lütkepohl (1991: 40-1)

Let us make a final comment by revisiting instantaneous causality. Using the prediction error MA representation following ARMA(p,q) (3.3.27a), we conclude from Theorem 3.3 that provided that the covariance matrix

$$\begin{bmatrix} \sigma_{11,\epsilon} & \sigma_{12,\epsilon} \\ \sigma_{21,\epsilon} & \sigma_{22,\epsilon} \end{bmatrix}$$

is nonsingular,
there is no instantaneous causality between \( x_{1t} \) and \( x_{2t} \) if and only if
\[ E(\varepsilon_{1t}\varepsilon_{2t}) = 0. \]

The conditions so far derived of Granger-causality are of stable bivariate processes. In the next chapter, unstable bivariate ARMA processes are considered. We also need to recall the concept "integration" because many nonstationary processes become stationary after differencing.
CHAPTER 4

COINTEGRATION


There are many time series (for example, in economics) which are not stationary, but some of them after being differenced a specific number (say d) of times, they can at least be approximated by stationary processes. In a bivariate setup $x_{1t}$ and $x_{2t}$ may be differenced different numbers of times $d_1$ and $d_2$ before each is stationary. (The numbers $d$, $d_1$ and $d_2$ are known as orders of integration). It was pointed out for the first time in Granger (1981) that a vector of time series, all of which are stationary only after differencing, may have linear combinations which are stationary without differencing. Such variables are said to be cointegrated.

This chapter is based on those bivariate nonstationary ARMA processes, which are going to be stationary after differencing where the two components require same order, say $d$, of integration. They are called ARIMA process as defined in Chapter 3. Also, we said they are integrated processes. Our study is about those components whose some linear function is not nonstationary any more.

4.1 COINTEGRATED PROCESSES


Some variables may have an equilibrium of some kind, for example in economics supply to the market, and demand by the consumer. That is, there are often "common" influences on (two economic) variables. As a result the two processes are never "too" different. At the same time these influences do not necessarily constrain the variables to be stationary, they may be very nonstationary. But because these variables are not too different, some linear function of them will be stationary.

The following definition is presented in Engle and Granger (1987: 253),
Lütkepohl (1991: 352), to name a few.

**Definition (Cointegrated time series)**

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate nonstationary time series. Then the series is said to be cointegrated of order \((d,b)\), denoted CI\((d,b)\) where \(b: 0 < b \leq d\) if:

1. \( x_{1t} \sim I(d), \quad x_{2t} \sim I(d) \).
2. There exists a vector \( \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \) such that the linear combination
   \[
   \alpha^T x_t = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \alpha_1 x_{1t} + \alpha_2 x_{2t}
   \]
   is integrated of order \(d-b\), \(I(d-b)\).
3. \(b\) is the largest integer such that (2) is possible.
4. \( \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \), which is not unique, is called a cointegrating vector.

Our focus will be mainly on \(d=b=1\). In general, as Lütkepohl (1991: 351) would say, \( z_t = \alpha^T x_t = \alpha_1 x_{1t} + \alpha_2 x_{2t} \) represents a random deviation from equilibrium, and this equilibrium as (almost) achieved when \( (z_t \approx 0) \) \( z_t = 0 \).

Engle and Granger (1987: 251) do emphasize the fact that each of these time series may be wandering extensively, but once they are cointegrated they wander as a group. The two components of a cointegrated time series share some common nonstationary components or "common trends" and hence, they tend to have certain similar movements in their long-term behaviour.

Before we focus on examples we discuss the given definition of cointegration. The assumption is that we start with two time series which are integrated of same order. Also, linear combinations are not necessarily stationary, but compared to original components the linear combination is of reduced integrating order \(d-b\). Lastly, any multiple \( k\alpha, (k\neq 0) \), is a cointegrating vector for any cointegrating vector \( \alpha \).
The following example is given by Hallman & Kamstra (1989: 190) and we want to 'spot' its features, but our emphasis is to check if it is cointegrated or not. To investigate cointegration we follow a method given by Lütkepohl (1991: 354) which is based on SVD.

**Example 4.1**
Suppose the model is given by

\[ x_{1t} = x_{1,t-1} + \varepsilon_{1t} \]  

(4.1.1a)

\[ x_{2t} = x_{1t} + \varepsilon_{2t} \]  

(4.1.1b)

which are known to be a random walk plus noise model which is nonstationary. In fact, the random walk is the nonstationary part of the model. Also,

\[ \Delta x_{1t} = \varepsilon_{1t} \]

is a white noise model and is stationary. Then

\[ x_{1t} \sim I(1). \]

Also,

\[ \Delta x_{2t} = \varepsilon_{1t} + \varepsilon_{2t} - \varepsilon_{2,t-1} = \varepsilon_{t} \]

is also white noise.

Therefore

\[ x_{1t} \sim I(1), \quad x_{2t} \sim I(1). \]

We present (4.1.1) in matrix notation as

\[
\begin{bmatrix}
1-L & 0 \\
-L & 1
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

which is bivariate AR(1) in form, or bivariate IAR(1,1).

To mention in passing, \( \phi_{12} = 0 \) implies from "characterization of Granger-causality" that \( x_{2t} \) does not Granger-cause \( x_{1t} \) while \( \phi_{21} = -1 \) (\( \phi_{21} \neq 0 \)) implies that \( x_{1t} \) Granger-causes \( x_{2t} \).

Using Lütkepohl (1991: 354)'s approach:
Eigenvector corresponding to $\lambda_1 = 2$:

$$\Pi \Pi^t [a] = 2[a]$$

$$\begin{bmatrix} a \\ b \end{bmatrix}$$

$a = 0, \ b = b$ is any real number.

That is, the unit length eigenvector associated with $\lambda_1 = 2$ is:

$$u = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Let

$$\alpha = \sqrt{\lambda_1} = \sqrt{2}$$

The method suggest that $v'$ is a cointegrating vector, where

$$v = \alpha^{-1} \Pi' u$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Therefore, a cointegrating vector is
In the following example we use a bivariate structural model, namely the random walk plus noise to illustrate the concept of integration using "common trends" approach of Harvey (1989: 450) and dynamic factor analysis Harvey (1989: 449-452) and Lütkepohl (1991: 422-423).

Example 4.2
The random walk plus noise process is given by

\[
\begin{bmatrix}
  x_{1t}
  \\
  x_{2t}
\end{bmatrix}
= \begin{bmatrix}
  \mu_{1t}
  \\
  \mu_{2t}
\end{bmatrix} + \begin{bmatrix}
  \epsilon_{1t}
  \\
  \epsilon_{2t}
\end{bmatrix}
\]

(4.1.2a)

\[
\begin{bmatrix}
  \mu_{1t}
  \\
  \mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
  \mu_{1,t-1}
  \\
  \mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
  \eta_{1t}
  \\
  \eta_{2t}
\end{bmatrix}
\]

(4.1.2b)

**Common trends approach**
Harvey (1989: 452 (8.5.5))
Here we assume that the trend is such that \( \mu_{1t} \) and \( \mu_{2t} \) share some common factor \( \mu_t \). By assigning (in the stated equation of Harvey) \( \beta = 0, \bar{\mu} = 0 \) then (4.1.2) becomes

\[
\begin{bmatrix}
  x_{1t}
  \\
  x_{2t}
\end{bmatrix}
= \begin{bmatrix}
  1
  \\
  0
\end{bmatrix} \mu_t + \begin{bmatrix}
  \epsilon_{1t}
  \\
  \epsilon_{2t}
\end{bmatrix}
\]

(4.1.2c)

\[
\mu_t = \mu_{t-1} + \eta_t
\]

(4.1.2d)

Now, \( \mu_t \) is a random walk and is nonstationary, and

\[\Delta \mu_t = \eta_t \] is stationary.

Then, because they are linear on \( \mu_t, x_{1t} \) and \( x_{2t} \) are both nonstationary.

Further,

\[\Delta x_{1t} = \eta_t + \epsilon_{1t} - \epsilon_{1,t-1} \] is stationary,
and
\[ \Delta x_{2t} = \theta \eta_t + \varepsilon_{2t} - \varepsilon_{2,t-1} \] is also stationary.

Thus
\[ x_{1t} \sim I(1) \text{ and } x_{2t} \sim I(1). \]

By noting that \( \mu_t \) is the (only) nonstationary component of (4.1.2c) and its loading vector is \( [1 \ 0]^T \), by removing \( \mu_t \) by premultiplying (4.1.2c) by \( [\theta \ -1]^T \) then
\[
\begin{bmatrix}
\theta \\
-1
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\theta \\
1
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]
is nonstationary (or I(0)).

A cointegrating vector is
\[
\begin{bmatrix}
\theta \\
-1
\end{bmatrix}
\]

**Dynamic factor analysis approach**


Because the random walk has two components \( \mu_{1t} \) and \( \mu_{2t} \), we assume that
\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
\]
depends on \( 1 < 2 \) unobserved common factor \( f_t \) and on factors
\[
\begin{bmatrix}
u_{1t} \\
v_{2t}
\end{bmatrix}
\]
(cf. (13.2.18) Lütkepohl (1991: 423) and taking \( \mu_t = f_t \))

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix}
\begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix}
\]

where \( \begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix} \) is a vector of factor loadings \( \theta_1 \neq 0, \theta_2 \neq 0 \), and \( \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix} \) is a white noise process. Substituting in (4.1.2a) we obtain
and substituting in the bivariate random walk (4.1.2b) we obtain

\[
\begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix}
\begin{bmatrix}
\mu_t \\
\mu_{t-1}
\end{bmatrix}
+ \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

We note that, as \( \mu_{1t} \) and \( \mu_{2t} \) are (nonstationary) random walks, \( \mu_t \) is nonstationary. Hence, \( x_{1t} \) and \( x_{2t} \) are nonstationary, but

\[
\begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix}
\begin{bmatrix}
\Delta \mu_t \\
\Delta \mu_{t-1}
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^*
\end{bmatrix}
\]

is a stationary process. From (4.1.2e)

\[
\Delta \begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\theta_1 \\
\theta_2
\end{bmatrix}
\Delta \begin{bmatrix}
\mu_t \\
\mu_{t-1}
\end{bmatrix}
+ \Delta \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^*
\end{bmatrix}
+ \Delta \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
= \begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
\]

That is

\( \Delta x_{1t} = v_{1t} \) and \( \Delta x_{2t} = v_{2t} \)

are white noise stationary processes. Therefore, \( x_{1t} \sim I(1) \) and \( x_{2t} \sim I(1) \).

By premultiplying (4.1.2e) using \([-\theta_2 \theta_1]\), then

\[
\begin{bmatrix}
-\theta_2 \\
\theta_1
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
-\theta_2 \\
\theta_1
\end{bmatrix}
\begin{bmatrix}
u_{1t} + \epsilon_{1t} \\
u_{2t} + \epsilon_{2t}
\end{bmatrix}
\]

is a stationary (white noise) process, that is, \( I(0) \).
Therefore a cointegrating vector is $[-\theta_2 \ \theta_1]$.

Therefore, with any of the above approaches

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \sim CI(1,1).$$

### 4.2 Cointegrated Bivariate ARMA Processes

We state without proofs two theorems which will be used in the forthcoming results, definitions and discussions. Proofs of these theorems are given in Appendix A.

**Theorem 4.1 (Singular Value Decomposition (SVD))**

Aoki (1987: 70)

An $m \times n$ real matrix $A$ of rank $r$ can be written as

$$A = U\Sigma V' = U_1 \Sigma_r V_1$$

where

$$U'U = I_m, \quad V'V = I_n$$

and

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_r = \text{diag}(\sigma_1, \sigma_2, ..., \sigma_r)$$

$$U = \begin{bmatrix} U_1 & U_2 \\ (m \times m) & (m \times (m \times r)) \end{bmatrix}, \quad V = \begin{bmatrix} V_1 & V_{n-r} \\ (n \times m) & (n \times (n \times r)) \end{bmatrix}$$

and the (singular) values are such that $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$.

The matrix $U_1$ is constructed by using orthonormal eigenvectors with (positive) eigenvectors of $AA'$ in decreasing order, and

$$V_1 = \Sigma^{-1}_r U_1' A$$

For proof of SVD see Theorem A.1 in Appendix A.
Theorem 4.2 (Spectral Decomposition Representation (SDR))

Aoki (1987: 253)

Suppose an \( nxn \) matrix \( A \) has \( n \) linearly independent eigenvectors \( u_1, \ldots, u_n \) and \( n \) eigenvalues \( \lambda_1 > \lambda_2 > \ldots > \lambda_n \). Define \( U = [u_1, \ldots, u_n] \), then the columns of \( U \) are linearly independent so that \( U^{-1} \) exists. Let \( V = U^{-1} \) with rows \( v_i' \), \( i = 1, \ldots, n \). Then \( u_i \) is the right eigenvector and \( v_i' \) the left eigenvector of \( \lambda_i \) because

\[
v_i' A = \lambda_i v_i', \quad A u_i = \lambda_i u_i
\]

Then by construction we have

\[
AU = U\Lambda, \quad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)
\]  

(4.2.2a)

Upon postmultiplying (4.2.2a) by \( V \) we obtain

\[
A = UAV = \sum_{i=1}^{n} \lambda_i u_i v_i'
\]

(4.2.2b)

Proof is outlined in Theorem A.3 of Appendix A.

A property from (4.2.2b), whose short proof is by induction, is

\[
A^k = U\Lambda^k V = \sum_{i=1}^{n} \lambda_i^k u_i v_i
\]

(4.2.2c)

Proof

\( k = 2 \)

\[
A^2 = (U\Lambda V)(U\Lambda V)
\]

\[
= U\Lambda^2 V : VU = I, \quad V = U^{-1}
\]

We assume that (4.2.2c) is true for \( k-1 \), that is assume

\[
A^{k-1} = U\Lambda^{k-1} V
\]

Then

\[
A^k = A^{k-1}A
\]

\[
= (U\Lambda^{k-1} V)(U\Lambda V)
\]

\[
= U\Lambda^k V : VU = I, \quad \Lambda^{k-1} \Delta = \Lambda^k
\]

which completes the proof.
Reinsel (1993: 43)

Let us discuss cointegration of a nonstationary bivariate AR(1) model,

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} =
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
z_{1,t-1} \\
z_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\tag{4.2.3}
\]

where suppose that \( \Phi \) has one eigenvalue equal to one, and the other \( |\lambda| < 1 \). From SDR there is a nonsingular 2×2 matrix \( U = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \), each \( u_i : 2 \times 1 \) such that

\[
\begin{bmatrix}
v_1' \\
v_2'
\end{bmatrix}
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix} =
\begin{bmatrix}
1 & 0 \\
0 & \lambda
\end{bmatrix}
\]

where

\[
V = U^{-1} = \begin{bmatrix} v_1' \\ v_2' \end{bmatrix}
\]

Define

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} =
\begin{bmatrix}
v_1' \\
v_2'
\end{bmatrix}
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = V \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix}
\]

and

\[
\begin{bmatrix}
a_{1t} \\
a_{2t}
\end{bmatrix} = V \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},
\]

then, premultiplication of (4.2.3) by \( V \)

\[
V \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} =
\begin{bmatrix}
v_1' \\
v_2'
\end{bmatrix}
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix}
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} + V \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\]

or

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 \\
0 & \lambda
\end{bmatrix}
\begin{bmatrix}
z_{1,t-1} \\
z_{2,t-1}
\end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}
\]
Thus
\[ z_{1t} = z_{1,t-1} + a_{1t} \]
is a nonstationary random walk, while
\[ z_{2t} = \lambda z_{2,t-1} + a_{2t} \]
is a stationary AR(1) process.

Since
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = U \begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \begin{bmatrix}
  u_1 & u_2
\end{bmatrix} \begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix}
\]
\[ = u_1 z_{1t} + u_2 z_{2t} \]

then this bivariate AR(1) process is a linear combination of a nonstationary (random walk) component and a stationary (AR(1)) component \( z_{2t} \).

Conversely, \( z_{2t} = v_2' x_{1t} \) is a linear combination of components of the original nonstationary bivariate series \( [x_{1t}, x_{2t}] \) that is stationary, and hence
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix}, \text{ (upon assuming that } z_{1t} \sim I(1), z_{2t} \sim I(1)) \text{, is a cointegrated process with } v_2 \text{ as cointegrating vector.}
\]

An extension to ARMA(1,1) of previous argument about AR(1) follows:

Reinsel (1993: 42)
Let us assume that the following bivariate ARMA is I(1), where
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \begin{bmatrix}
  \phi_{11} & \phi_{12} \\
  \phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
  z_{1,t-1} \\
  z_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix} + \begin{bmatrix}
  \theta_{11} & \theta_{12} \\
  \theta_{21} & \theta_{22}
\end{bmatrix} \begin{bmatrix}
  \varepsilon_{1,t-1} \\
  \varepsilon_{2,t-1}
\end{bmatrix} \tag{4.2.4}
\]
and all components are as in (4.2.3) and we define
We have once more, a nonstationary random walk

\[(I-L)z_{1t} = v_{1t}\]

and a stationary AR(1) model

\[(1-\lambda L)z_{2t} = v_{2t}\]

Also,

\[
\begin{bmatrix}
\alpha_{1t} \\
\alpha_{2t}
\end{bmatrix} = \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{21} & \theta_{22}
\end{bmatrix} \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

then

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1}
\end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t}
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t}
\end{bmatrix} - \begin{bmatrix} a_{1,t-1}^* \\ a_{2,t-1}^*
\end{bmatrix}
\]

We have once more, a nonstationary random walk

\[(I-L)z_{1t} = v_{1t}\]

and a stationary AR(1) model

\[(1-\lambda L)z_{2t} = v_{2t}\]

Also,

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = U \begin{bmatrix} z_{1t} \\ z_{2t}
\end{bmatrix} = \begin{bmatrix} u_{1} \end{bmatrix} z_{1t} + \begin{bmatrix} u_{2} \end{bmatrix} z_{2t}
\]

is a nonstationary bivariate ARMA(1,1) model which is a linear combination of a nonstationary component and a stationary one. The stationary component is again given by

\[
z_{2t} = v_{2} \begin{bmatrix} x_{1t} \\ x_{2t}
\end{bmatrix}
\]

which shows that we have a cointegrated process with cointegrating vector \(v_{2}\).

Suppose now we have a nonstationary ARMA(p,q) where each \(z_{it} \sim I(1)\),

\[
\sum_{i=0}^{p} \begin{bmatrix} \phi_{11,i}L^i & \phi_{12,i}L^i \\
\phi_{21,i}L^i & \phi_{22,i}L^i
\end{bmatrix} \begin{bmatrix} x_{1t} \\
x_{2t}
\end{bmatrix} = \sum_{j=0}^{q} \begin{bmatrix} \theta_{11,j}L^j & \theta_{12,j}L^j \\
\theta_{21,j}L^j & \theta_{22,j}L^j
\end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]  \hspace{1cm} (4.2.5)
where
\[
\begin{bmatrix}
\phi_{11,0} & \phi_{12,0} \\
\phi_{21,0} & \phi_{22,0}
\end{bmatrix}
= \begin{bmatrix}
\theta_{11,0} & \theta_{12,0} \\
\theta_{21,0} & \theta_{22,0}
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}.
\]

**Definition (Cointegrating rank)**
A process given by (4.2.5) is said to be cointegrated of rank \( r \leq 2 \) if
\[
\Phi(1) = \sum_{i=0}^{p} \begin{bmatrix}
\phi_{11,i} & \phi_{12,i} \\
\phi_{21,i} & \phi_{22,i}
\end{bmatrix}
\text{ has rank } r > 0.
\]

We note that the AR matrix has been written conveniently for discussion. \((I_2 + \Phi_1 + \ldots \) instead of usual \( I_2 - \Phi_1 - \ldots \)).

From the above definition, cointegration is possible only for \( r=1 \) and \( r=2 \). Previous examples were examples of \( r=1 \) case. From Corollary A.2 of Appendix A, Lütkepohl (1991: 354-5) and Reinsel (1993: 42-3) if \( \Phi(1) \) has rank \( r \in \{1, 2\} \) then there exists \( 2 \times r \) matrices \( H \) and \( C \) of rank \( r \) each such that \( \Phi(1) = HC \).

**Definition (Cointegrating matrix, loading matrix)**
Lütkepohl (1991: 355)
The matrix \( C \) defined above is a cointegrating matrix or a matrix of cointegrating vectors, and \( H \) is (sometimes) called a loading matrix.

The MA matrix and the error terms do not play a role in the cointegration as seen from the above definition. We disregard this part by defining:
\[
\begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
= \sum_{j=0}^{q} \begin{bmatrix}
\theta_{11,j} & \theta_{12,j} \\
\theta_{21,j} & \theta_{22,j}
\end{bmatrix}
L^j \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}
\]
and we write (4.2.5) in shorthand as
\[
x_t = -\Phi_1 x_{t-1} - \ldots - \Phi_p x_{t-p} + v_t
\] (4.2.6)
By assumption we said $x_{1t} \sim I(1)$ and $x_{2t} \sim I(1)$, then

\[
\begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix} \sim I(0).
\]

Lütkepohl (1991: 355)

We write (4.2.6) as (a stationary process):

\[
\Delta z_t = \Phi(1)z_{t-1} + F_1 \Delta z_{t-1} + \cdots + F_{p-1} \Delta z_{t-p+1} + v_t
\]

where

\[
F_i = \Phi_{i+1} + \cdots + \Phi_p, \quad i = 1, \ldots, p-1
\]

and rearranging we have

\[
\Phi(1)z_{t-1} = \Delta z_t - F_1 \Delta z_{t-1} - \cdots - F_{p-1} \Delta z_{t-p+1} - v_t
\]

which is stationary because all terms in the right hand side of the equation are stationary. That is $\Phi(1)z_{t-1} = Hz_{t-1}$ is stationary, and $(H' H)^{-1} H \Phi(1)z_{t-1} = Cz_{t-1}$ is also stationary. This implies that each row of $C$ is a cointegrating vector. If $r=2$ there are two linearly independent cointegrating vectors, and if $r=1$ there is only one.

We note that if rank $\Phi(1) = 0$, then $\Phi(1) = 0$, and this is a reason for $r>0$ in the definition of cointegrating rank. The following remark is made from this comment.

**Remark**

Suppose (4.2.5) is such that $x_{1t} \sim I(d), x_{2t} \sim I(d), d>0$. Then $z_t$ is cointegrated if and only if $\Phi(1) \neq O_d$.

### 4.3 Granger-Causal Cointegrated Bivariate ARMA Processes


We consider a unique bivariate ARMA(p,q) process given by
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \sum_{i=1}^p \begin{bmatrix}
  \phi_{11,i} & \phi_{12,i}
\end{bmatrix} \begin{bmatrix}
  x_{1,t-i} \\
  x_{2,t-i}
\end{bmatrix} + \sum_{i=0}^q \begin{bmatrix}
  \theta_{11,i} & \theta_{12,i} \\
  \theta_{21,i} & \theta_{22,i}
\end{bmatrix} \begin{bmatrix}
  \varepsilon_{1,t-i} \\
  \varepsilon_{2,t-i}
\end{bmatrix}
\]

(4.3.1)

where
\[
\begin{bmatrix}
  \theta_{11,0} & \theta_{12,0} \\
  \theta_{21,0} & \theta_{22,0}
\end{bmatrix} = \begin{bmatrix}
  1 & 0 \\
  0 & 1
\end{bmatrix}
\]

Lütkepohl (1991: 229) gives the optimal forecasts
\[
\begin{aligned}
\begin{bmatrix}
  z_{1t}(\ell) \\
  z_{2t}(\ell)
\end{bmatrix} &= \begin{cases}
  \sum_{i=1}^p \begin{bmatrix}
  \phi_{11,i} & \phi_{12,i}
\end{bmatrix} \begin{bmatrix}
  z_{1t}(\ell-i) \\
  z_{2t}(\ell-i)
\end{bmatrix} + \sum_{j=1}^q \begin{bmatrix}
  \theta_{11,j} & \theta_{12,j} \\
  \theta_{21,j} & \theta_{22,j}
\end{bmatrix} \begin{bmatrix}
  \varepsilon_{1,t+\ell-j} \\
  \varepsilon_{2,t+\ell-j}
\end{bmatrix}, & h \leq q \\
  \sum_{i=1}^p \begin{bmatrix}
  \phi_{11,i} & \phi_{12,i}
\end{bmatrix} \begin{bmatrix}
  z_{1t}(\ell-i) \\
  z_{2t}(\ell-i)
\end{bmatrix}, & h > q
\end{cases}
\end{aligned}
\]

and
\[
\begin{bmatrix}
  z_{1t}(j) \\
  z_{2t}(j)
\end{bmatrix} = \begin{bmatrix}
  x_{1,t+j} \\
  x_{2,t+j}
\end{bmatrix} \text{ for } j \leq 0.
\]

The 1-step ahead forecast for \( z_{1t} \), based on the information set
\[
\begin{bmatrix}
  z_{1s} \\
  z_{2s}
\end{bmatrix} \text{ for } s \leq t
\]

\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \begin{bmatrix}
  1 & 0
\end{bmatrix} \begin{bmatrix}
  z_{1t}(1) \\
  z_{2t}(1)
\end{bmatrix}
\]
\begin{align*}
&= \sum_{i=1}^{p} \phi_{11,i} x_{1,t+1-i} + \sum_{i=1}^{p} \phi_{12,i} x_{2,t+1-i} + \sum_{i=1}^{q} \theta_{11,i} \varepsilon_{1,t+1-i} \\
&\quad + \sum_{i=1}^{q} \theta_{12,i} \varepsilon_{2,t+1-i} \\
&= \varepsilon_{1,t+1}
\end{align*}

The forecast error is

\begin{align*}
&z_{1,t+1} - x_{1t} = \begin{bmatrix}
\{ x_{1s} \}_{s \leq t} \\
\{ x_{2s} \}
\end{bmatrix}
\begin{bmatrix} 1 \ 0 \end{bmatrix} - \begin{bmatrix} 1 \ 0 \end{bmatrix} \begin{bmatrix} x_{1,t+1} \\
\varepsilon_{1,(t+1)}
\end{bmatrix} \\
&= \varepsilon_{1,t+1}
\end{align*}

If \( x_{1t} \) has a univariate ARMA(p,q) form, then

\begin{align*}
x_{1t} &= \sum_{i=1}^{p} \psi_{1,i} x_{t-i} + \sum_{i=0}^{q} \psi_{2,i} \eta_{t-i}, \quad \psi_{1,0} = 1 \\
&= \sum_{i=1}^{p} \phi_{11,i} x_{1,t-i} + \sum_{i=1}^{p} \phi_{12,i} x_{2,t+1-i} + \sum_{i=0}^{q} \theta_{11,i} \varepsilon_{1,t-i} \\
&\quad + \sum_{i=1}^{q} \theta_{12,i} \varepsilon_{2,t-i}
\end{align*}

and

\begin{align*}
x_{1t} = \begin{bmatrix} 1 \{ x_{1s} \}_{s \leq t} \\
\{ x_{2s} \}
\end{bmatrix} = \sum_{i=1}^{p} \psi_{1,i} x_{1,t+1-i} + \sum_{i=1}^{q} \psi_{2,i} \eta_{t+1-i} \\
&= \psi_{1,0} x_{1,t+1} \\
&= \psi_{1,0} x_{1,t+1} + \sum_{i=1}^{q} \psi_{2,i} \eta_{t+1-i}
\end{align*}

The forecast error is

\begin{align*}
&z_{1,t+1} - x_{1t} = \begin{bmatrix} 1 \{ x_{1s} \}_{s \leq t} \\
\{ x_{2s} \}
\end{bmatrix} = \eta_{t+1}
\end{align*}

The predictors (4.3.2) and (4.3.3) are identical if and only if:
\[ \varepsilon_{1t} = \eta_t \text{ for all } t. \]

Equality of predictors is therefore equivalent to \( x_{1t} \) having the ARMA form

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= \sum_{i=1}^{p} \begin{bmatrix}
    \psi_{1,i} & 0 \\
    \psi_{2,i} & 0
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-i} \\
    x_{2,t-i}
\end{bmatrix}
+ \sum_{j=0}^{q} \begin{bmatrix}
    \psi_{1,j} & 0 \\
    \psi_{2,j} & 0
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_{1,t-j} \\
    \varepsilon_{2,t-j}
\end{bmatrix}
\]

\[
= \sum_{i=1}^{p} \begin{bmatrix}
    \phi_{11,i} & \phi_{12,i} \\
    \phi_{21,i} & \phi_{22,i}
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-i} \\
    x_{2,t-i}
\end{bmatrix}
+ \sum_{j=0}^{q} \begin{bmatrix}
    \theta_{11,j} & \theta_{12,j} \\
    \theta_{21,j} & \theta_{22,j}
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_{1,t-j} \\
    \varepsilon_{2,t-j}
\end{bmatrix}
\]

and therefore uniqueness of ARMA implies that

\[
\psi_{1,i} = \phi_{11,i}, \quad \psi_{1,j} = 0, \quad i = 1, \ldots, p
\]

\[
\psi_{2,j} = \theta_{11,j}, \quad \psi_{2,j} = 0, \quad j = 1, \ldots, q
\]

We have proved the next theorem

**Theorem 4.4 (Granger-noncausality)**

Lütkepohl (1991: 375-378)

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate ARMA(p,q) given by (4.3.1) with unique parameters. Then

\[
x_{1t}(\{x_{1s} \mid s \leq t\}) = x_{1t}(\{x_{1s} \mid s \leq t\})
\]

if and only if

\[
\phi_{12,i} = \theta_{12,j} = 0 \text{ for } i = 1, \ldots, p, \quad j = 1, \ldots, q
\]

From the \( \ell \)-step forecast expression (4.3.1) it is evident that equality of 1-step forecasts implies equality of the \( \ell \)-step forecasts for \( \ell = 2, 3, \ldots \), especially on noting that \( \ell \)-step forecast is 1-step forecasts for
The following corollary is based on Theorem 4.4 and the above argument. The proof is analogous to the one of Corollary 3.2.1.

Corollary 4.4.1
Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate ARMA(p,q) given by (4.3.1) with unique parameters. Then

\[
\begin{bmatrix} x_{1t}(l-1) \\ x_{2t}(l-1) \end{bmatrix}
\]

\[
\begin{bmatrix} x_{1t}(l|\{x_{1s}|s \leq t\}) \\ x_{2t}(l|\{x_{2s}|s \leq t\}) \end{bmatrix}
\]

if and only if

\[ \phi_{12,i} = 0 \quad \text{for all} \quad i = 1, \ldots, p, \quad j = 1, \ldots, q \]

The above discussion and results were not restricted to any specific ARMA, except that parameters are unique for each process we consider, hence it does agree even for cointegrated processes. We have the following corollary, the proof is similar to that of Corollary 3.2.2.

Corollary 4.4.2
Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate ARMA(p,q) as in (4.3.1) with unique parameters (where the process may be cointegrated). Then

\( x_{2t} \) does not Granger-cause \( x_{1t} \)

if and only if

\[ \phi_{12,i} = 0 \quad \text{for all} \quad i, j \]
To forecast is to predict or to estimate future values which we call forecasts. Once certified acceptable, forecasts may be used to develop other forecasts. This is a reason for striving for accurate forecasts, as well as flexibility of forecasts. This chapter focuses on forms and behavior of point forecasts and forecast regions for the models we will consider. It is through these that we study and gain valuable insight to certain properties of these models. Our tools are the SSR and the Kalman forecasts.

We recall from Chapter 2 that the SSR is given by

\[ z_t = Wz_{t-1} + Bv_t, \quad BG^1 = 0 \tag{5.0.1a} \]
\[ z_t = Hz_t + g v_t \tag{5.0.1b} \]

where

\[ E(v_t) = 0, \quad E(v_t v_t^\prime) = I \]

We have derived the following estimates which are also given in these references.

Given \( z(t|t) \) and \( \Sigma_z(t|t) \).

The \( \ell \)-step ahead (Kalman) point forecast is

\[ z(t+\ell|t) = HW^\ell z(t|t) \tag{5.0.2} \]

and the corresponding forecast MSE is
The bivariate IMA(1,1) process has the form

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= \begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix} + \begin{bmatrix}
    \theta_{11} & \theta_{12} \\
    \theta_{21} & \theta_{22}
\end{bmatrix} \begin{bmatrix}
    \varepsilon_{1,t-1} \\
    \varepsilon_{2,t-1}
\end{bmatrix}
\]

(5.1.1)

5.1 BIVARIATE IMA(1,1) PROCESS

Let

\[
\begin{bmatrix}
    z_{11}^{(1)} \\
    z_{21}^{(1)} \\
    \vdots \\
    z_{11}^{(2)} \\
    z_{21}^{(2)}
\end{bmatrix} = \begin{bmatrix}
    x_{1t} \\
    x_{2t} \\
    \vdots \\
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    \varepsilon_{1,t-1} + \theta_{11} \varepsilon_{1,t-1} + \theta_{12} \varepsilon_{2,t-1} + \varepsilon_{1t} \\
    \varepsilon_{2,t-1} + \theta_{21} \varepsilon_{1,t-1} + \theta_{22} \varepsilon_{2,t-1} + \varepsilon_{2t}
\end{bmatrix}
\]

(5.0.3)

That is,
Now we define
\[ \epsilon_t = \Sigma_{\epsilon \epsilon}^{\frac{1}{2}} v_t \]
where
\[ E(v_t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \text{cov}(v_t) = I_2 \]
and also set
\[ H = \begin{bmatrix} I_2 & O_2 \end{bmatrix}, \quad B_{11} = B_{21} = \Sigma_{\epsilon \epsilon}^{\frac{1}{2}} \]
then
\[
\begin{bmatrix}
  z_{t}^{(1)} \\
  \vdots \\
  z_{t}^{(2)} 
\end{bmatrix}
= \begin{bmatrix}
  I_2 & \Theta \\
  O_2 & O_2 
\end{bmatrix}
\begin{bmatrix}
  z_{t-1}^{(1)} \\
  \vdots \\
  z_{t-1}^{(2)} 
\end{bmatrix}
+ \begin{bmatrix}
  B_{11} \\
  B_{21} 
\end{bmatrix} v_t
\]
\[
\begin{bmatrix}
  x_{1t} \\
  x_{2t} 
\end{bmatrix}
= \begin{bmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 
\end{bmatrix}
\begin{bmatrix}
  z_{t}^{(1)} \\
  \vdots \\
  z_{t}^{(2)} 
\end{bmatrix}
\]

5.1.1 Point forecasts

Given
\[ z(t|t) = \begin{bmatrix} z_1(t|t) \\ z_2(t|t) \end{bmatrix} = \begin{bmatrix} z(t|t) \\ \epsilon(t|t) \end{bmatrix} \]
then the \( \ell \)-step ahead forecast (\( \ell \in \mathbb{N} \)) are
\[ z(t+\ell|t) = HW^\ell z(t|t) \]
where
\[ W = \begin{bmatrix} I_2 & \Theta \\ O_2 & O_2 \end{bmatrix}, \quad W^2 = W \cdot W = \begin{bmatrix} I_2 & \Theta \\ O_2 & O_2 \end{bmatrix} \]
That is,
\[ W^2 = W \]
We derive \( W^\ell \) by induction, where \( \ell=1 \), and \( \ell=2 \) are valid from previous step.
Suppose (the induction step of) \( t-1 \) case is such that 
\[ W^{t-1} = W. \]
Then case \( t \) is
\[ W^t = W^{t-1}W = W \cdot W : \text{using induction step} \]
\[ = W^2 = W \text{ for all } \ell \in \mathbb{N} \]

\[ HW^\ell = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} I_2 & \Theta \\ 0 & O \end{bmatrix} \]
\[ = \begin{bmatrix} I_2 & \Theta \end{bmatrix} \text{ for all } \ell \in \mathbb{N}. \]

We define \( W^0 = I. \)

The \( \ell \)-step forecast for any \( \ell \in \mathbb{N} \) is
\[
\begin{bmatrix} z_{1t}(\ell) \\ z_{2t}(\ell) \end{bmatrix} = HW^\ell \begin{bmatrix} z_{11}(t|t) \\ z_{21}(t|t) \end{bmatrix}
= [I \quad \Theta] \begin{bmatrix} z_{11}(t|t) \\ z_{21}(t|t) \end{bmatrix}
= z(t|t) + \Theta \epsilon(t|t)
\]
\[
\begin{bmatrix} z_1(t|t) \\ z_2(t|t) \end{bmatrix} = \begin{bmatrix} \theta_{11} \epsilon_1(t|t) + \theta_{12} \epsilon_2(t|t) \\ \theta_{21} \epsilon_1(t|t) + \theta_{22} \epsilon_2(t|t) \end{bmatrix}
\]

5.1.2 Forecast (or minimum) mean square error matrices (MMSE)

Given \( \Sigma_{zz}(t|t) \), we denote \( \Sigma_{zz}(\ell) = \Sigma_{zz}(t+\ell|t) \). Using (5.0.3)

\[
\Sigma_{xx}(\ell) = HW^\ell \Sigma_{zz}(t|t)(HW^\ell)' + \sum_{j=0}^{\ell-1} HW^j B (HW^j B)' + GG'
\]
and the partition
\[
\Sigma_{zz}(t|t) = \begin{bmatrix}
\Sigma_{xx}(t|t) & \Sigma_{xz}(t|t) \\
\Sigma_{ez}(t|t) & \Sigma_{ee}(t|t)
\end{bmatrix}
\]

then
\[
\Sigma_{xx}(1) = \begin{bmatrix} I & \Theta \end{bmatrix} \Sigma_{zz}(t|t) \begin{bmatrix} I_2 \\ \Theta \end{bmatrix} + HBB'H' + GG'
\]

We have \( GG' = 0 \) and
\[
BB' = \begin{bmatrix} B_{11} & B'_{11} \\ B_{21} \\
\end{bmatrix}
\]
\[
= \begin{bmatrix} \Sigma_{ee} & \Sigma_{ee} \\ \Sigma_{ee} & \Sigma_{ee} \\
\end{bmatrix}
\]
\[
HBB'H' = \begin{bmatrix} I_2 & O_2 \end{bmatrix} BB' \begin{bmatrix} I_2 \\ O_2 \end{bmatrix} = \Sigma_{ee}
\]

Substituting and simplifying:
\[
\Sigma_{xx}(1) = \Sigma_{xx}(t|t) + \Theta \Sigma_{ez}(t|t) + \Sigma_{xz}(t|t) \Theta' + \Theta \Sigma_{ee}(t|t) \Theta' + \Sigma_{ee} 
\]
\[(5.1.3)\]

For \( \ell \in \mathbb{N}, \ell \geq 2 \) from (5.0.3) with \( GG' = 0 \),
\[
\Sigma_{zz}(t) = H W^\ell \Sigma_{zz}(t|t)(H W^\ell)' + HBB'H' + \sum_{j=1}^{\ell-1} H W^j BB'(H W^j)'
\]
\[
H W^\ell \Sigma_{zz}(t|t)(H W^\ell)' = H W \Sigma_{zz}(t|t)(H W)'
\]
\[
= \begin{bmatrix} I & \Theta \end{bmatrix} \Sigma_{zz}(t|t) \begin{bmatrix} I \\ \Theta \end{bmatrix}
\]
\[
= \Sigma_{xx}(1) \quad \text{: from } \ell=1
\]
\[ \begin{bmatrix} \sigma_{11,\ell}^{(1)} & \sigma_{12,\ell}^{(1)} \\ \sigma_{21,\ell}^{(1)} & \sigma_{22,\ell}^{(1)} \end{bmatrix} \]

\[ \text{HBB}'H' = \Sigma_{\varepsilon\varepsilon} \text{ from } \ell=1 \]

and lastly, with \( W^j = W \) for all \( j \in \mathbb{W} \),

\[ \text{HW}^j \text{BB}'(\text{HW}^j)' = \text{HWBB}'(\text{HW})' \]

\[ = [I \quad \Theta] \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\varepsilon} \\ \Sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\varepsilon} \end{bmatrix} \begin{bmatrix} I \\ \Theta' \end{bmatrix} \]

\[ = (I + \Theta)\Sigma_{\varepsilon\varepsilon}(I + \Theta)' \]

Then

\[ \Sigma_{xx}(\ell) = \Sigma_{xx}(1) + \sum_{j=1}^{\ell-1} (I + \Theta)\Sigma_{\varepsilon\varepsilon}(I + \Theta)' \]

\[ = \Sigma_{xx}(1) + (\ell-1)(I + \Theta)\Sigma_{\varepsilon\varepsilon}(I + \Theta)' \]

\[ = \Sigma_{xx}(1) + (\ell-1)(I + \Theta)(I + \Theta)' : \text{ if } \Sigma_{\varepsilon\varepsilon} = I_2 \]

\[ = \begin{bmatrix} \sigma_{11,\ell}^{(1)} & \sigma_{12,\ell}^{(1)} \\ \sigma_{21,\ell}^{(1)} & \sigma_{22,\ell}^{(1)} \end{bmatrix} + (\ell-1) \begin{bmatrix} 1+\theta_{11} & \theta_{12} \\ \theta_{21} & 1+\theta_{22} \end{bmatrix} \begin{bmatrix} 1+\theta_{11} & \theta_{12} \\ \theta_{12} & 1+\theta_{22} \end{bmatrix} \]

and \( \Sigma_{xx}(\ell) \) is a linear function of \( \ell \). The assumption \( \Sigma_{\varepsilon\varepsilon} = I_2 \) has been made because \( \Sigma_{\varepsilon\varepsilon} \) has no role in describing SUTSE and hence no effect on the concept but it provided convenience. Expanding we have
5.1.3 Granger-causal IMA(1,1)

The general bivariate IMA(1,1) process given by (5.1.1) is a feedback system between \(x_{1t}\) and \(x_{2t}\). We now restrict this to \(x_{1t}\) not Granger-causal for \(x_{2t}\) so that \(\theta_{21} = 0\), but \(x_{2t}\) still Granger-causes \(x_{1t}\). Then in (5.1.1) \(\Theta\) is replaced by

\[
\Theta = \begin{bmatrix}
\theta_{11} & \theta_{12} \\
0 & \theta_{22}
\end{bmatrix}
\]

The point forecast (5.1.2) becomes

\[
\begin{bmatrix}
x_{1t}(t) \\
x_{2t}(t)
\end{bmatrix} = \begin{bmatrix}
x_1(t|t) \\
x_2(t|t)
\end{bmatrix} + \begin{bmatrix}
\theta_{11}\varepsilon_1(t|t) + \theta_{12}\varepsilon_2(t|t) \\
\theta_{22}\varepsilon_2(t|t)
\end{bmatrix}
\]

and even in point forecasts there is absolutely no contribution by \(x_{1t}\) in the prediction of \(x_{2t}\) (that is, even \(\varepsilon_1(t|t)\) is not involved), whereas the point forecasts of \(x_{1t}\) are "Granger-caused" by \(x_{2t}\) only through the error term \(\varepsilon_2(t|t)\), which further emphasizes SUTSE.

In (5.1.3) and (5.1.4) there is some "reduction" in the factors \(\text{cov}(\varepsilon_{1t}, \varepsilon_{2t})\) because \(\theta_{21} = 0\). Then \(\Sigma_{xx}(1)\) is affected at \(\sigma_{12,x}(1) = \sigma_{21,x}(1)\) and \(\sigma_{22,x}(1)\) so that the new MSE may be denoted \(\sigma_{1j,x}(1)\) at the affected moments.

Lastly, the MMSE at (5.1.4) is reduced to

\[
\Sigma_{xx}(t) = \begin{bmatrix}
\sigma_{11,x}(1) & \sigma_{12,x}(1) \\
\sigma_{21,x}(1) & \sigma_{22,x}(1)
\end{bmatrix}
\]

\[
+ (t-1)\begin{bmatrix}
(1+\theta_{11})^2 + \theta_{12}^2 \\
\theta_{21}(1+\theta_{11}) + \theta_{12}(1+\theta_{22}) \\
\theta_{21}(1+\theta_{11}) + \theta_{12}(1+\theta_{22}) \\
\theta_{21}^2 + (1+\theta_{22})^2
\end{bmatrix}
\]

(5.1.4)
\[ \Sigma_{xx}(\ell) = \begin{bmatrix} \sigma_{11, x(1)}^* & \sigma_{12, x(1)}^* \\ \sigma_{21, x(1)}^* & \sigma_{22, x(1)}^* \end{bmatrix} + (\ell-1) \begin{bmatrix} (1+\theta_{11})^2 + \theta_{12}^2 & \theta_{12}(1+\theta_{22}) \\ \theta_{12}(1+\theta_{22}) & (1+\theta_{22})^2 \end{bmatrix} \]

Even though the MMSE at (5.1.6) is still unbounded, it has been reduced because the contribution of \( x_{1t} \) is reduced.

**5.2 BIVARIATE IMA(2,2)**

The IMA(2,2) process has been shown in Chapter 2 to be

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = 2 \begin{bmatrix} x_{1, t-1} \\ x_{2, t-1} \end{bmatrix} - \begin{bmatrix} x_{1, t-2} \\ x_{2, t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11, 1} & \theta_{12, 1} \\ \theta_{21, 1} & \theta_{22, 1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1, t-1} \\ \varepsilon_{2, t-1} \end{bmatrix} \\
+ \begin{bmatrix} \theta_{11, 2} & \theta_{12, 2} \\ \theta_{21, 2} & \theta_{22, 2} \end{bmatrix} \begin{bmatrix} \varepsilon_{1, t-2} \\ \varepsilon_{2, t-2} \end{bmatrix}
\]

and we write in short-hand as

\[
x_t = 2x_{t-1} - x_{t-2} + \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \Theta_2 \varepsilon_{t-2}
\]

The SSR follows, where we take \( v_t \) to be of zero mean vector and covariance matrix \( I_2 \), then let

\[
z_t = \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ \vdots \\ z_t^{(3)} \\ z_t^{(4)} \end{bmatrix} = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}
\]
\[
\begin{bmatrix}
2I_2 & -I_2 & \Theta_1 & \Theta_2 \\
I_2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & I & 0 \\
\end{bmatrix}
\begin{bmatrix}
z_{t-1}^{(1)} \\
z_{t-1}^{(2)} \\
z_{t-1}^{(3)} \\
z_{t-1}^{(4)} \\
\end{bmatrix}
+ \begin{bmatrix}
\Sigma^{\frac{1}{2}} \varepsilon \\
\end{bmatrix}
\begin{bmatrix}
v_t \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
W_{11} & W_{12} \\
0 & W_{22} \\
\end{bmatrix}
z_{t-1} + Bv_t
\]

\[
x_t = \begin{bmatrix}
x_{1t} \\
x_{2t} \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
I_2 & 0 & 0 & 0 \\
\end{bmatrix}z_t
\]

\[
= Hz_t
\]

### 5.2.1 Point forecasts

We first note the following result and definitions

\[
W_{22}^2 = \begin{bmatrix}
O & O \\
I_2 & O \\
\end{bmatrix}
\begin{bmatrix}
O & O \\
I_2 & O \\
\end{bmatrix} = O_4
\]

We define

\[
W^0 = I_8 \quad \text{and} \quad W^0_{11} = I_4
\]

Now we derive expression for \( W^\ell \) by induction

\[
W = \begin{bmatrix}
W_{11} & W_{12} \\
0 & W_{22} \\
\end{bmatrix}
\]

\[
W^2 = \begin{bmatrix}
W_{11} & W_{12} \\
0 & W_{22} \\
\end{bmatrix}
\begin{bmatrix}
W_{11} & W_{12} \\
0 & W_{22} \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
O & O \\
I_2 & O \\
\end{bmatrix}
\begin{bmatrix}
O & O \\
I_2 & O \\
\end{bmatrix}
\begin{bmatrix}
O & O \\
I_2 & O \\
\end{bmatrix}
\]
For induction step, suppose $\ell - 1$ is satisfied, that is

$$W^{\ell - 1} = \begin{bmatrix} W_{11}^{\ell - 1} & W_{11}^{\ell - 2}W_{12} + W_{11}^{\ell - 3}W_{12}W_{22} \\ 0 & 0 \end{bmatrix}$$

then if the above is true, then for $\ell \geq 2$ ($\ell \in \mathbb{N}$)

$$W^{\ell} = W^{\ell - 1}W$$

$$= \begin{bmatrix} W_{11}^{\ell} & W_{11}^{\ell - 1}W_{12} + W_{11}^{\ell - 2}W_{12}W_{22} \\ 0 & 0 \end{bmatrix}$$

Also,

$$HW = \begin{bmatrix} 2I_2 & -I_2 & \Theta_1 & \Theta_2 \end{bmatrix}$$

so that given

$$z(t|t) = \begin{bmatrix} z(t|t) \\ x(t-1|t) \\ \varepsilon(t|t) \\ \varepsilon(t-1|t) \end{bmatrix}$$

then the 1-step ahead forecast using (5.0.2) is

$$\begin{bmatrix} x_{1t}(1) \\ x_{2t}(1) \end{bmatrix} = HWz(t|t)$$

$$= 2z(t|t) - z(t-1|t) + \Theta_1\varepsilon(t|t) + \Theta_2\varepsilon(t-1|t)$$

(5.2.1)
For the \( \ell \)-step forecast \((\ell \in \mathbb{N}, \ell \geq 2)\) we need the \( W^j \) in a more informative form. From

\[
W^j = \begin{bmatrix}
W^j_{11} & W^j_{12} \sum_{t=1}^{j-1} W^j_{12} W^j_{22} \\
0 & 0
\end{bmatrix}, \quad j \geq 2, \quad j \in \mathbb{N}
\]

We derive expression for \( W^k_{11} \) by induction, \( k \in \mathbb{N} \).

\[
W^1_{11} = \begin{bmatrix} 2I & -I \\ I & 0 \end{bmatrix}
\]

\[
W^2_{11} = W^1_{11} W^1_{11} = \begin{bmatrix} 3I & -2I \\ 2I & -I \end{bmatrix}
\]

\[
W^3_{11} = W^2_{11} W^1_{11} = \begin{bmatrix} 4I & -3I \\ 3I & -2I \end{bmatrix}
\]

For induction step, suppose

\[
W^{k-1}_{11} = \begin{bmatrix} kI & -(k-1)I \\ (k-1)I & -(k-2)I \end{bmatrix}
\]

then using the above, we have

\[
W^k_{11} = W^{k-1}_{11} W^1_{11} = \begin{bmatrix} (k+1)I & -kI \\ kI & -(k-1)I \end{bmatrix}, \quad k \geq 2, \quad k \in \mathbb{N}
\]
\[ W_{11}^k W_{12} = \begin{bmatrix} (k+1)I & -kI \\ kI & -(k-1)I \end{bmatrix} \begin{bmatrix} \Theta_1 & \Theta_2 \\ 0 & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} (k+1)\Theta_1 & (k+1)\Theta_2 \\ k\Theta_1 & k\Theta_2 \end{bmatrix} \]

\[ W_{11}^k W_{12} W_{22} = \begin{bmatrix} (k+1)\Theta_1 & (k+1)\Theta_2 \\ k\Theta_1 & k\Theta_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ I_2 & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} (k+1)\Theta_2 & 0 \\ k\Theta_2 & 0 \end{bmatrix} \]

\[ W_{11}^{k-1} W_{12} + W_{11}^{k-2} W_{12} W_{22} = \begin{bmatrix} k\Theta_1 & k\Theta_2 \\ (k-1)\Theta_1 & (k-1)\Theta_2 \end{bmatrix} + \begin{bmatrix} (k-1)\Theta_2 & 0 \\ (k-2)\Theta_2 & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} k(\Theta_1 + \Theta_2) - \Theta_2 & k\Theta_2 \\ (k-1)(\Theta_1 + \Theta_2) - \Theta_2 & (k-1)\Theta_2 \end{bmatrix} \]

Therefore, for \( \ell \in \mathbb{N}, \ell \geq 2 \)

\[ W^\ell = \begin{bmatrix} (\ell+1)I_2 & -\ell I_2 & \ell(\Theta_1 + \Theta_2) - \Theta_2 & \ell \Theta_2 \\ \ell I_2 & -(\ell-1)I_2 & (\ell-1)(\Theta_1 + \Theta_2) & (\ell-1)\Theta_2 \\ 0_4 & 0_4 \end{bmatrix} \]

\[ HW^\ell = \begin{bmatrix} (\ell+1)I_2 & -\ell I_2 & \ell(\Theta_1 + \Theta_2) - \Theta_2 & \ell \Theta_2 \\ \ell I_2 & -(\ell-1)I_2 & (\ell-1)(\Theta_1 + \Theta_2) & (\ell-1)\Theta_2 \end{bmatrix}, \quad \ell \geq 2 \]

Then the \( \ell \)-step ahead forecast is given by

\[ z(\ell) = HW^\ell z(t|t) \]

\[ = (\ell+1)z(t|t) - \ell z(t-1|t) + \ell(\Theta_1 + \Theta_2)\epsilon(t|t) - \Theta_2 \epsilon(t|t) \]

\[ + \ell \Theta_2 \epsilon(t-1|t), \quad \ell \geq 2 \]

(5.2.2)
In long-hand this forecast is

\[
\begin{bmatrix}
    x_{1t}(\ell) \\
    x_{2t}(\ell)
\end{bmatrix}
= (\ell+1) \begin{bmatrix}
    x_1(t|t) \\
    x_2(t|t)
\end{bmatrix}
- \ell \begin{bmatrix}
    x_1(t-1|t) \\
    x_2(t-1|t)
\end{bmatrix}
+ \ell \begin{bmatrix}
    \theta_{11,1} + \theta_{11,2} & \theta_{12,1} + \theta_{12,2} \\
    \theta_{21,1} + \theta_{21,2} & \theta_{22,1} + \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_1(t|t) \\
    \varepsilon_2(t|t)
\end{bmatrix}
- \begin{bmatrix}
    \theta_{11,2} & \theta_{12,2} \\
    \theta_{21,2} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_1(t|t) \\
    \varepsilon_2(t|t)
\end{bmatrix}
\]

(5.2.2*)

This forecast is an unbounded function of \( \ell \), and the rate of expansion is high. Since the first two terms of the equation may be written as

\[
\begin{bmatrix}
    x_1(t|t) \\
    x_2(t|t)
\end{bmatrix}
+ \ell \left[ \begin{bmatrix}
    x_1(t|t) \\
    x_2(t|t)
\end{bmatrix}
- \begin{bmatrix}
    x_1(t-1|t) \\
    x_2(t-1|t)
\end{bmatrix} \right]
\]

the estimate of slope is

\[
\begin{bmatrix}
    \hat{s}_{1t} \\
    \hat{s}_{2t}
\end{bmatrix}
= \begin{bmatrix}
    x_1(t|t) \\
    x_2(t|t)
\end{bmatrix}
- \begin{bmatrix}
    x_1(t-1|t) \\
    x_2(t-1|t)
\end{bmatrix}
+ \begin{bmatrix}
    \theta_{11,1} + \theta_{11,2} & \theta_{12,1} + \theta_{12,2} \\
    \theta_{21,1} + \theta_{21,2} & \theta_{22,1} + \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_1(t|t) \\
    \varepsilon_2(t|t)
\end{bmatrix}
- \begin{bmatrix}
    \theta_{11,2} & \theta_{12,2} \\
    \theta_{21,2} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
    \varepsilon_1(t|t) \\
    \varepsilon_2(t|t)
\end{bmatrix}
\]

which controls forecasts through the signs (negative, positive or zero) and when zero it does not affect them in any way.
5.2.2 **MMSE matrices**

Let $\Sigma_{zz}(t|t)$ be given which for convenience we present with partition based on numbering on $z_t$ vectors as:

$$
\Sigma_{zz}(t|t) = \begin{bmatrix}
\Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\
\Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\
\Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\
\Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44}
\end{bmatrix}
$$

The MMSE for 1-step forecast (5.2.1) is

$$
\Sigma_{zz}(1) = HW\Sigma_{zz}(t|t)(HW)' + HBB'H'
$$

where

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

so that

$$
HBB'H' = \Sigma_{\varepsilon\varepsilon}
$$

and

$$
HW\Sigma_{zz}(t|t)(HW)' = \begin{bmatrix}
2I & -I & \Theta_1 & \Theta_2 \end{bmatrix} \Sigma_{zz}(t|t) \begin{bmatrix} 2I \\
-I \\
\Theta_1' \\
\Theta_2'
\end{bmatrix} = 4\Sigma_{11} + \Sigma_{22} + \Theta_1\Sigma_{33}\Theta_1' + \Theta_2\Sigma_{44}\Theta_2' - 2\Sigma_{21}
$$

$$
-2\Sigma_{12} + 2\Theta_1\Sigma_{31} - \Theta_1\Sigma_{32} + 2\Theta_2\Sigma_{41} - \Theta_2\Sigma_{42}
$$

$$
+ 2\Sigma_{13}\Theta_1' - \Sigma_{23}\Theta_1' + 2\Sigma_{14}\Theta_2' - \Sigma_{24}\Theta_2' + \Theta_1'\Sigma_{34}\Theta_1' + \Theta_1'\Sigma_{34}\Theta_1'
$$

$$
+ \Theta_1'\Sigma_{34}\Theta_1'
$$
and for \( \ell \in \mathbb{N}, \ell \geq 2 \) the \( \ell \)-step ahead MMSE forecast matrices are given by

\[
\Sigma_{xx}(\ell) = \text{HW}^\ell \Sigma_{zz}(t|t)(\text{HW}^\ell)' + \sum_{j=0}^{\ell-1} \text{HW}^j \text{BB}^j (W^j)' H'
\]

where

\[
S_1 = \text{HW}^\ell \Sigma_{zz}(t|t)(\text{HW}^\ell)'
\]

\[
= [(\ell+1)I - \ell I \quad \ell(\Theta_1 + \Theta_2) - \Theta_2 \quad \ell \Theta_2] \Sigma_{zz}(t|t)[(\ell+1)I - \ell I \quad \ell(\Theta_1 + \Theta_2) - \Theta_2 \quad \ell \Theta_2]'
\]

\[
= (\ell+1)^2 \Sigma_{11} + \ell^2 \Sigma_{22} + \ell^2 (\Theta_1 + \Theta_2) \Sigma_{33}(\Theta_1 + \Theta_2)'
+ \ell^2 \Theta_2 \Sigma_{44} \Theta_2' + \ell^2 \Theta_2 \Sigma_{32} + \ell (\ell+1) \Sigma_{13}(\Theta_1 + \Theta_2)'
+ \ell (\ell+1) \Sigma_{14} \Theta_2' + \ell (\ell+1) (\Theta_1 + \Theta_2) \Sigma_{31} + \ell (\ell+1) \Theta_2 \Sigma_{41}
- \ell^2 \Sigma_{23}(\Theta_1 + \Theta_2)'
+ \ell \Sigma_{23} \Theta_2' + \ell^2 (\Theta_1 + \Theta_2) \Sigma_{34} \Theta_2'
+ \ell^2 \Theta_2 \Sigma_{43}(\Theta_1 + \Theta_2)'
+ \Theta_2 \Sigma_{33} \Theta_2'
- \ell (\ell+1) (\Sigma_{12} + \Sigma_{21}) - (\ell+1) (\Sigma_{13} \Theta_2' + \Theta_2 \Sigma_{31})
- \ell \Theta_2 (\Sigma_{34} + \Sigma_{43}) \Theta_2'
- \ell^2 \Theta_2 \Sigma_{42} - \ell^2 (\Theta_1 + \Theta_2) \Sigma_{32}
- \ell (\Theta_1 + \Theta_2) \Sigma_{33} \Theta_2'
- \ell \Theta_2 \Sigma_{33}(\Theta_1 + \Theta_2)'
\]

and

\[
S_2 = \sum_{j=0}^{\ell-1} \text{HW}^j \text{BB}^j (\text{HW}^j)'
\]

\[
= \text{BB}^\ell H' + \sum_{j=1}^{\ell-1} \text{HW}^j \text{BB}(\text{HW}^j)' H'
\]

where

\[
\text{BB}^\ell H' = \Sigma_{\epsilon \epsilon} \quad \text{as before}
\]

and

\[
\text{HW}^j = (j+1) \Sigma_{\epsilon \epsilon}^\frac{1}{2} \quad j(\Theta_1 + \Theta_2) \Sigma_{\epsilon \epsilon}^\frac{1}{2} - \Theta_2 \Sigma_{\epsilon \epsilon}^\frac{1}{2}
\]
so that

\[ \text{HW}^j \text{B}(\text{HW}^j \text{B})' = (j+1)^2 \sum_{\epsilon \epsilon} + j(j+1)(\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} - (j+1) \Theta_2 \sum_{\epsilon \epsilon} \]

\[ + j(j+1) \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' + \int^2 (\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' \]

\[ - j \Theta_2 \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' - (j+1) \sum_{\epsilon \epsilon} \Theta_2' - j(\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} \Theta_2' \]

\[ + \Theta_2 \sum_{\epsilon \epsilon} \Theta_2' \]

We have

\[ \ell \sum_{j=1}^{\ell-1} (j+1)^2 = \frac{\ell(\ell+1)(2\ell+1)}{6} - 1 \]

\[ \ell \sum_{j=1}^{\ell-1} j(j+1) = \frac{\ell(\ell-1)(\ell+1)}{3} \]

\[ \ell \sum_{j=1}^{\ell-1} j^2 = \frac{\ell(\ell-1)(2\ell-1)}{6} \]

\[ \ell \sum_{j=1}^{\ell-1} j = \frac{\ell(\ell-1)}{2} \]

and

\[ \ell \sum_{j=1}^{\ell-1} (j+1) = \frac{\ell(\ell+1)}{2} - 1 \]

Then

\[ S_2 = \sum_{\epsilon \epsilon} + \left\{ \frac{\ell(\ell+1)(2\ell+1)}{6} - 1 \right\} \sum_{\epsilon \epsilon} + \frac{\ell(\ell-1)(\ell+1)}{3} (\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} \]

\[ - \left\{ \frac{\ell(\ell+1)}{2} - 1 \right\} \Theta_2 \sum_{\epsilon \epsilon} + \frac{\ell(\ell-1)(\ell+1)}{3} \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' \]

\[ + \frac{\ell(\ell-1)(2\ell-1)}{6} (\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' - \frac{\ell(\ell-1)}{2} \Theta_2 \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' \]

\[ - \left\{ \frac{\ell(\ell+1)}{2} - 1 \right\} \sum_{\epsilon \epsilon} \Theta_2' - \frac{\ell(\ell-1)}{2} (\Theta_1 + \Theta_2) \sum_{\epsilon \epsilon} (\Theta_1 + \Theta_2)' \]

\[ + (\ell-1) \Theta_2 \sum_{\epsilon \epsilon} \Theta_2' \]
\[
= \frac{\ell (\ell+1)(2\ell+1)}{6} \Sigma_{\varepsilon \varepsilon} + \frac{\ell (\ell-1)(\ell+1)}{3} (\theta'_{1} \Sigma_{\varepsilon \varepsilon} + \Sigma_{\varepsilon \varepsilon} \theta'_{1}) \\
+ \frac{(\ell-2)(\ell-1)(2\ell+3)}{6} (\theta'_{2} \Sigma_{\varepsilon \varepsilon} + \Sigma_{\varepsilon \varepsilon} \theta'_{2}) + \frac{\ell (\ell-2)(\ell-4)}{6} (\theta'_{2} \Sigma_{\varepsilon \varepsilon} \theta'_{1}) \\
+ \frac{(\ell-2)(\ell-4)}{6} (\ell-1) \theta'_{2} \Sigma_{\varepsilon \varepsilon} \theta'_{2} + \frac{\ell (\ell-1)(\ell-2)}{6} (\theta'_{1} \Sigma_{\varepsilon \varepsilon} \theta'_{1} + \theta'_{1} \Sigma_{\varepsilon \varepsilon} \theta'_{2})
\]

Thus, this MMSE "expands" far more than for the IMA(1,1)

5.2.3 Granger-causal IMA(2,2)

We assume that \( x_{1t} \) does not Granger-cause \( x_{2t} \), but \( x_{2t} \) is Granger-causal for \( x_{1t} \), then from the result we had in Granger-causality, \( \theta_{21,1} = \theta_{21,2} = 0 \). Then the 1-step ahead forecast (5.2.1) becomes

\[
\begin{bmatrix}
    x_{1t}(1) \\
    x_{2t}(1)
\end{bmatrix} = \begin{bmatrix}
    2x_{1}(t|t) - x_{1}(t-1|t) + [\theta_{11,1}\varepsilon_{1}(t|t) + \theta_{12,1}\varepsilon_{2}(t|t)] \\
    x_{2}(t|t) - x_{2}(t-1|t) + \theta_{21,2}\varepsilon_{2}(t-1|t)
\end{bmatrix}
\]

and the \( \ell \)-step ahead forecast (5.2.2*) becomes

\[
\begin{bmatrix}
    x_{1t}(\ell) \\
    x_{2t}(\ell)
\end{bmatrix} = \begin{bmatrix}
    x_{1}(t|t) + \ell [x_{1}(t|t) - x_{1}(t-1|t)] \\
    x_{2}(t|t) - x_{2}(t-1|t)
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
    (\theta_{11,1} + \theta_{11,2})\varepsilon_{1}(t|t) + (\theta_{12,1} + \theta_{12,2})\varepsilon_{2}(t|t) \\
    \theta_{22,1}\varepsilon_{2}(t-1|t)
\end{bmatrix}
\]

\[
- \begin{bmatrix}
    \theta_{11,2}\varepsilon_{1}(t|t) + \theta_{12,2}\varepsilon_{2}(t|t) \\
    \theta_{22,2}\varepsilon_{2}(t|t)
\end{bmatrix}
\]

\[
+ \ell \begin{bmatrix}
    \theta_{11,2}\varepsilon_{1}(t-1|t) + \theta_{12,2}\varepsilon_{2}(t-1|t) \\
    \theta_{22,2}\varepsilon_{2}(t-1|t)
\end{bmatrix}
\]
In these forecasts the behaviour SUTSE is still reflected, and also from the 
\( \varepsilon(t|t) \) we can read off that \( x_{1t} \) does not Granger-cause \( x_{2t} \), but the 
reverse is true that \( x_{2t} \) Granger-causes \( x_{1t} \), and can also be read off in the 
same way.

The MMSE matrix (5.2.4) for the \( \ell \)-step ahead forecast takes same form, but 
in \( \Theta_1 \) and \( \Theta_2 \), the elements \( \theta_{21,1} = \theta_{21,2} = 0 \) will reflect that \( x_{1t} \) does 
not Granger cause \( x_{2t} \).

5.3 BIVARIATE SUTSE ARIMA (1,1,2) MODEL

The bivariate ARIMA(1,1,2) model is of the form

\[
\begin{bmatrix}
1 - \frac{\phi_{11}}{\phi_{12}}L & \frac{\phi_{12}}{\phi_{12}}L \\
\frac{\phi_{21}}{\phi_{22}} & \frac{\phi_{22}}{\phi_{22}}
\end{bmatrix}
\begin{bmatrix}
I - L
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix} + \begin{bmatrix}
\theta_{11,1} & \theta_{12,1} \\
\theta_{21,1} & \theta_{22,1}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-1} \\
\varepsilon_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\theta_{11,2} & \theta_{12,2} \\
\theta_{21,2} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-2} \\
\varepsilon_{2,t-2}
\end{bmatrix}
\]

which is SUTSE when \( \phi_{12} = \phi_{21} = 0 \), and then becomes

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
1 + \phi_{11} & 0 \\
0 & 1 + \phi_{22}
\end{bmatrix}
\begin{bmatrix}
x_{1,t-1} \\
x_{2,t-1}
\end{bmatrix} - \begin{bmatrix}
\phi_{11} & 0 \\
0 & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
x_{1,t-2} \\
x_{2,t-2}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1,t} \\
\varepsilon_{2,t}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
\theta_{11,1} & \theta_{12,1} \\
\theta_{21,1} & \theta_{22,1}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-1} \\
\varepsilon_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\theta_{11,2} & \theta_{12,2} \\
\theta_{21,2} & \theta_{22,2}
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1,t-2} \\
\varepsilon_{2,t-2}
\end{bmatrix}
\]

We will use

\[
\Phi = \begin{bmatrix}
\phi_{11} & 0 \\
0 & \phi_{22}
\end{bmatrix}
\]
assume $|\phi_{11}|, |\phi_{22}| < 1$ and define $\Phi^0 = I_2$ so that

$$(I - \Phi)^{-1} = \sum_{j=0}^{\infty} \Phi^j$$

and

$$I + \Phi + \ldots + \Phi^k = (I - \Phi)^{-1}(I - \Phi^{k+1})$$

$$= (I - \Phi^{k+1})(I - \Phi)^{-1}$$

Define

$$z_t = \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ \vdots \\ z_t^{(3)} \\ z_t^{(4)} \end{bmatrix} = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-1} \\ \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}$$

$$= \begin{bmatrix} I + \Phi & -\Phi & \Theta_1 & \Theta_2 \\ I & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ z_t^{(3)} \\ z_t^{(4)} \end{bmatrix} + \begin{bmatrix} \Sigma_{1/2} \varepsilon \varepsilon \\ \Sigma_{1/2} \varepsilon \varepsilon \end{bmatrix} v_t$$

$$z_t = \begin{bmatrix} I_2 & 0 & 0 & 0 \end{bmatrix} z_t$$

We use induction to derive the $W^k$ matrices.

$$W = \begin{bmatrix} I + \Phi & -\Phi & \Theta_1 & \Theta_2 \\ I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}$$
\[ W^2 = \begin{bmatrix} (I+\Phi)^2 - \Phi & -(I+\Phi)\Phi & (I+\Phi)\Theta_1+\Theta_2 & (I+\Phi)\Theta_2 \\ I+\Phi & -\Phi & \Theta_1 & \Theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]
\[ = \begin{bmatrix} I+\Phi + \Phi^2 & -(I+\Phi)\Phi & (I+\Phi)\Theta_1+\Theta_2 & (I+\Phi)\Theta_2 \\ I+\Phi & -\Phi & \Theta_1 & \Theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ W^3 = W^2 W = \begin{bmatrix} \sum_{j=0}^{3} \Phi^j & -\sum_{j=0}^{2} \Phi^j \Phi & \sum_{j=0}^{2} \Phi^j \Theta_1 + \sum_{j=0}^{1} \Phi^j \Theta_2 & \sum_{j=0}^{2} \Phi^j \Theta_2 \\ \sum_{j=0}^{2} \Phi^j & -\sum_{j=0}^{1} \Phi^j \Phi & \sum_{j=0}^{1} \Phi^j \Theta_1 + \Theta_2 & \sum_{j=0}^{1} \Phi^j \Theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

Then we assume that the induction step is true. That is, assume

\[ W^{\ell-1} = \begin{bmatrix} \sum_{j=0}^{\ell-1} \Phi^j & -\sum_{j=0}^{\ell-2} \Phi^j \Phi & \sum_{j=0}^{\ell-2} \Phi^j \Theta_1 + \sum_{j=0}^{\ell-3} \Phi^j \Theta_2 & \sum_{j=0}^{\ell-2} \Phi^j \Theta_2 \\ \sum_{j=0}^{\ell-2} \Phi^j & -\sum_{j=0}^{\ell-3} \Phi^j \Phi & \sum_{j=0}^{\ell-3} \Phi^j \Theta_1 + \sum_{j=0}^{\ell-4} \Phi^j \Theta_2 & \sum_{j=0}^{\ell-3} \Phi^j \Theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

Then using the above step we obtain

\[ W^{\ell} = W^{\ell-1} W = \begin{bmatrix} \sum_{j=0}^{\ell} \Phi^j & -\sum_{j=0}^{\ell-1} \Phi^j \Phi & \sum_{j=0}^{\ell-1} \Phi^j \Theta_1 + \sum_{j=0}^{\ell-2} \Phi^j \Theta_2 & \sum_{j=0}^{\ell-1} \Phi^j \Theta_2 \\ \sum_{j=0}^{\ell-1} \Phi^j & -\sum_{j=0}^{\ell-2} \Phi^j \Phi & \sum_{j=0}^{\ell-2} \Phi^j \Theta_1 + \sum_{j=0}^{\ell-3} \Phi^j \Theta_2 & \sum_{j=0}^{\ell-2} \Phi^j \Theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ \ell \in \mathbb{N}, \quad \ell \geq 3 \]
Suppose
\[
z(t|t) = \begin{bmatrix} x(t|t) \\ x(t-1|t) \\ \epsilon(t|t) \\ \epsilon(t-1|t) \end{bmatrix}
\] is given
\[
x(t-1|t) = c_1(t|t)
\]

\(t\)-step ahead forecasts are given by
\[
z(t+t|t) = HW^t z(t|t)
\]

\(t=1\)
\[
HW = \begin{bmatrix} I + \Phi & -\Phi & \Theta_1 & \Theta_2 \end{bmatrix}
\]
\[
\begin{bmatrix} z_{1t}(1) \\ z_{2t}(1) \end{bmatrix} = (I + \Phi)z(t|t) - \Phi z(t-1|t) + \Theta_1 \epsilon(t|t) + \Theta_2 \epsilon(t-1|t)
\]
\[
= \begin{bmatrix} 1+\phi_{11} & 0 \\ 0 & 1+\phi_{22} \end{bmatrix} \begin{bmatrix} z_1(t|t) \\ z_2(t|t) \end{bmatrix} - \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} z_1(t-1|t) \\ z_2(t-1|t) \end{bmatrix}
\]
\[
+ \begin{bmatrix} \theta_{11,1} & \theta_{12,1} \\ \theta_{21,1} & \theta_{22,1} \end{bmatrix} \begin{bmatrix} \epsilon_1(t|t) \\ \epsilon_2(t|t) \end{bmatrix} + \begin{bmatrix} \theta_{11,2} & \theta_{12,2} \\ \theta_{21,2} & \theta_{22,2} \end{bmatrix} \begin{bmatrix} \epsilon_1(t-1|t) \\ \epsilon_2(t-1|t) \end{bmatrix}
\]
which has SUTSE and feedback features \((\phi_{12} = \phi_{21} = 0; \theta_{12,i} \neq 0, \theta_{21,i} \neq 0)\)

\(t=2\)
\[
HW^2 = \begin{bmatrix} I + \Phi + \Phi^2 & -(I + \Phi)\Phi & (I + \Phi)\Theta_1 + \Theta_2 & (I + \Phi)\Theta_2 \end{bmatrix}
\]
so that
\[
\begin{bmatrix} z_{1t}(2) \\ z_{2t}(2) \end{bmatrix} = (I + \Phi + \Phi^2)z(t|t) - (I + \Phi)\Phi z(t-1|t) + (I + \Phi)\Theta_1 \epsilon(t|t) + \Theta_2 \epsilon(t-1|t) + (I + \Phi)\Theta_2 \epsilon(t-1|t)
\]
which also displays both SUTSE and feedback. Granger-causality is achieved by setting only one of $\theta_{12,1} = \theta_{12,2} = 0$ and $\theta_{21,1} = \theta_{21,2} = 0$.

for $\ell \geq 3$,

$$\text{HW}^\ell = (I - \Phi)^{-1} \left[ I - \Phi^\ell \Phi \cdots (I - \Phi^\ell)^\ell (I - \Phi^\ell)^{\ell - 1} \Theta_1 + (I - \Phi^\ell)^{\ell - 1} \Theta_2 + (I - \Phi^\ell)^{\ell} \Theta_2 \right]$$

$$\approx (I - \Phi)^{\ell - 1} \left[ I - \Phi \Theta_1 + \Theta_2 \Theta_2 \right] \text{ for very large } \ell$$

because $|\phi_{11}|, |\phi_{22}| < 1$

and

$$\Phi^k = \begin{bmatrix} \phi_{11}^k & 0 \\ 0 & \phi_{22}^k \end{bmatrix} \to O_2 \text{ as } k \to \infty$$

Then for very large horizon $\ell$, the forecast is

$$\begin{bmatrix} z_{1t}(\ell) \\ z_{2t}(\ell) \end{bmatrix} = (I - \Phi)^{-1} \left[ z(t|t) - \Phi z(t-1|t) + (\Theta_1 + \Theta_2) \varepsilon(t|t) + \Theta_2 \varepsilon(t-1|t) \right]$$
This forecast displays SUTSE and feedback behavior.

The forecast MSE will look almost like that of IMA(2,2), but expanding at a lower rate because of the damping AR matrix \( \Phi \) of the form:

\[
(I - \Phi)^{-1} = \begin{bmatrix}
(1 + \phi_{11})^{-1} & 0 \\
0 & (1 + \phi_{22})^{-1}
\end{bmatrix}
\]

The following section is a generalization of confidence intervals, but now carried out on bivariate processes jointly. The definition will be followed by analysis of three models, each of which is both Granger-causal and cointegrated. The models were chosen with the help of Prof. Markham.

5.4 FORECAST REGIONS

For convenience we assume that the models are under normality conditions, or can be approximated by normal random variables. The following definition is due to Johnson and Wichern (1992: 132), modified for forecasts after discussion with Prof. Markham.

**Definition (Forecast region)**

Let \( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \) be a bivariate process with \( \ell \)-step ahead forecasts given by \( \begin{bmatrix} x_{1t}(\ell) \\ x_{2t}(\ell) \end{bmatrix} \) and the corresponding MMSE given by \( \Sigma_{xx}(\ell) \), where \( \Sigma_{xx}(\ell) \) is
assumed to be nonsingular. A $100(1-\alpha)\%$ forecast region for \[
\begin{bmatrix}
    x_{1,t+\ell} \\
    x_{2,t+\ell}
\end{bmatrix}
\]
the set of all points inside the ellipse
\[
\begin{bmatrix}
    x_{1,t+\ell} - x_{1t}(\ell) \\
    x_{2,t+\ell} - x_{2t}(\ell)
\end{bmatrix}^\top
\Sigma_{xx}^{-1}(\ell)
\begin{bmatrix}
    x_{1,t+\ell} - x_{1t}(\ell) \\
    x_{2,t+\ell} - x_{2t}(\ell)
\end{bmatrix} = \chi^2_2(\alpha)
\] (5.4.1)
where $\chi^2_2(\alpha)$ is the upper $100\alpha\%$ probability of a chi-square variable of two degrees of freedom.

The ellipse is centered at \[
\begin{bmatrix}
    x_{1t}(\ell) \\
    x_{2t}(\ell)
\end{bmatrix}
\]
and the lengths of major and minor axes are $\chi^2_2(\alpha) \lambda_1$ and $\chi^2_2(\alpha) \lambda_2$ where $\lambda_1$ and $\lambda_2$ are distinct eigenvalues of the positive definite matrix $\Sigma_{xx}(\ell)$.

In all the following examples we focus on $95\%$ forecast regions so that $\chi^2_2(0.05) = 5.99$

Example 5.4.1
Consider the process
\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    0 & a \\
    0 & 1
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}, \ a \in \mathbb{R}
\] (5.4.2)
Initially, suppose $a \neq 0$.

$x_{2t}$ is a random walk, hence $x_{2t} \sim I(1)$ and $x_{2t}$ Granger-causes $x_{1t}$ because $\phi_{21} = a \neq 0$ (by assumption at the beginning).

As
\[
x_{1t} = ax_{2,t-1} + \varepsilon_{1t}, \ x_{1t} \sim I(1) \quad \text{as well.}
\]
\[
\Pi = I - \Phi
\]
has rank 1, hence of cointegrating rank 1. Let \( \alpha = \begin{bmatrix} -1 \\ a \end{bmatrix} \)

\[
\alpha' \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \alpha' \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \text{ is stationary (or I(0))}
\]

Therefore

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \sim CI(1,1)
\]

Let \( z(t|t) \) and \( \Sigma_{\varepsilon\varepsilon} = I \) be given, then considering \( z(t) \) as state vector of SSR, \( \ell \) step ahead forecasts are

\[
\begin{bmatrix} x_{1t}(\ell) \\ x_{2t}(\ell) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^{\ell} \begin{bmatrix} x_{1(t|t)} \\ x_{2(t|t)} \end{bmatrix} \quad \ell \geq 1 \ (\ell \in \mathbb{N})
\]

and MMSE matrices

\[
\begin{bmatrix} \sigma_{11,z}(\ell) & \sigma_{12,z}(\ell) \\ \sigma_{21,z}(\ell) & \sigma_{22,z}(\ell) \end{bmatrix} = \sum_{j=0}^{\ell-1} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^j \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}^j
\]

with \( W^0 \) defined as

\[
\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

Now

\[
W = \begin{bmatrix} 0 & a \\ 0 & 1 \end{bmatrix}
\]
\[ W^2 = WW = \begin{bmatrix} 0 & a \\ 0 & 1 \end{bmatrix} \]

Induction can be used to show that for \( t \in \mathbb{N}, \ t \geq 1 \)

\[ W^t = W \]

\[ WW^t = \begin{bmatrix} 0 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ a & 1 \end{bmatrix} = \begin{bmatrix} a^2 & a \\ a & 1 \end{bmatrix} \]

All point forecasts are then given by

\[
\begin{bmatrix} x_{1t}(t) \\ x_{2t}(t) \end{bmatrix} = \begin{bmatrix} 0 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t|t) \\ x_2(t|t) \end{bmatrix} = \begin{bmatrix} ax_2(t|t) \\ x_2(t|t) \end{bmatrix}
\]

and MMSE by

\[
\begin{bmatrix} \sigma_{11,x}(t) & \sigma_{12,x}(t) \\ \sigma_{21,x}(t) & \sigma_{22,x}(t) \end{bmatrix} = W^0(W')^0 + \sum_{j=1}^{t-1} W^j W^j \]

\[
= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (t-1) \begin{bmatrix} a^2 & a \\ a & 1 \end{bmatrix}
\]

Choosing \( a = 0.5 \), \( \mathbf{y}(t|t) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \) and \( t = 1, 2, 3 \)

\[
\begin{bmatrix} x_{1t}(1) \\ x_{2t}(1) \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}, \quad \Sigma_{xx}(1) = 1_2
\]

Using (5.4.1) the ellipse is

\[
\begin{bmatrix} x - 0.5 \\ y - 1 \end{bmatrix}' \begin{bmatrix} I^{-1} \\ I^{-1} \end{bmatrix} < \chi^2_2(\alpha) = c^2
\]

\[
\begin{bmatrix} y - 1 \\ y - 1 \end{bmatrix}
\]
Thus, the 95% confidence region is
\[(x - 0.5)^2 + (y-1)^2 < 5.99\]
a circle with centre \[\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}\] having radius \[\sqrt{5.99} = 2.447\].

That is, the 1-step ahead forecast region is the circle described.

\[
\begin{bmatrix}
  z_{1t}(2) \\
  z_{2t}(2)
\end{bmatrix} =
\begin{bmatrix}
  0.5 \\
  1
\end{bmatrix}, \quad 
\Sigma_{xx}(2) = 
\begin{bmatrix}
  1.25 & 0.5 \\
  0.5 & 2
\end{bmatrix}, \quad
v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \theta = 63.4^0
\]

eigenvalues of \(\Sigma_{xx}(2)\): \(\lambda^2 - 3.25\lambda + 2.25 = 0\)
\(\lambda_1 = 2.25, \quad \lambda_2 = 1\)

length of major axis: \(c \sqrt{\lambda_1} = \sqrt{2.25 \times 5.99} = 3.671\)
length of minor axis: \(c \sqrt{\lambda_2} = 2.447\)
centre \[\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}\]

From 1-step to 2-step only the major axis expanded while the minor coincides with the previous one of a circle.

\[
\begin{bmatrix}
  z_{1t}(3) \\
  z_{2t}(3)
\end{bmatrix} =
\begin{bmatrix}
  0.5 \\
  1
\end{bmatrix}, \quad 
\Sigma_{xx}(3) = 
\begin{bmatrix}
  1.5 & 1 \\
  1 & 3
\end{bmatrix}, \quad
v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \theta = 63.4^0
\]

eigenvalues of \(\Sigma_{xx}(3)\): \(\lambda^2 - 4.5\lambda + 3.5 = 0\)
\(\lambda_1 = 3.5, \quad \lambda_2 = 1\)

length of major axis: \(c \sqrt{\lambda_1} = 4.579\)
length of minor axis: \(c \sqrt{\lambda_2} = 2.447\)
centre \[\begin{bmatrix} 0.5 \\ 1 \end{bmatrix}\]
The expansion is seen once again only in the major axis, and the centre and direction of axis are the same.

Now if \( a = 0 \):

We would have had in (5.4.2) a white noise \( x_{1t} = \varepsilon_{1t} \) and a random walk \( x_{2t} = x_{2,t-1} + \varepsilon_{2t} \); no causality and because of different integration orders 0 and 1 there is no cointegration.

Point forecasts are

\[
\begin{bmatrix}
  x_{1t}(\ell) \\
  x_{2t}(\ell)
\end{bmatrix} =
\begin{bmatrix}
  0 \\
  z_2(t|t)
\end{bmatrix}
\]

and MMSE matrices are

\[
\Sigma_{x_{12}}(\ell) = \begin{bmatrix} 1 & 0 \\ 0 & \ell \end{bmatrix}
\]

\( \lambda_1 = \ell, \quad \lambda_2 = 1 \)

\( \nu_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \theta = 90^\circ \)

\( c\sqrt{\lambda_1} = \sqrt{5.99\ell} \), \( c\sqrt{\lambda_2} = \sqrt{5.99} \)

The centre is on the \( x_{2t} \) axis which serves also as major axis, and minor axis has fixed length \( c\sqrt{\lambda_2} = 2.447 \) and is parallel to the \( x_{1t} \) axis.

**Example 5.4.2**

Our next process is

\[
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} =
\begin{bmatrix}
  \phi_{11} & \phi_{12} \\
  0 & 1
\end{bmatrix}
\begin{bmatrix}
  x_{1,t-1} \\
  x_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix}, \quad |\phi_{11}| < 1, \quad \phi_{12} \in \mathbb{R}
\]

(5.4.3)

Assuming \( \phi_{11} \neq 0, \phi_{12} \neq 0 \), using the approach employed in the previous example,

\( x_{1t} \sim \text{I}(1), \quad x_{2t} \sim \text{I}(1) \)
and the process has cointegrating rank 1 so that

\[
\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix} \sim \text{CI}(1,1)
\]

With assumption $\phi_{12} \neq 0$ we have $\mathbf{x}_{1t}$ Granger-caused by $\mathbf{x}_{2t}$.

For forecasts we require

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1
\end{bmatrix}^2 = \begin{bmatrix}
\phi_{11}^2 & \phi_{11} \phi_{12} + \phi_{12} \\
0 & 1
\end{bmatrix}
\]

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1
\end{bmatrix}^3 = W^2W = \begin{bmatrix}
\phi_{11}^3 & \phi_{11}^2 \phi_{12} + \phi_{11} \phi_{12} \phi_{12} \\
0 & 1
\end{bmatrix}
\]

and we can show by induction that for $j \in \mathbb{N}$, $j \geq 2$

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1
\end{bmatrix}^j = \begin{bmatrix}
\phi_{11}^j & \sum_{k=0}^{j-1} \phi_{11}^{j-1-k} \phi_{12}^k \\
0 & 1
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0 & \phi_{12} \\
1-\phi_{11} & 1
\end{bmatrix}
\]

for large $j$

We define

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1
\end{bmatrix}^0 = I_2
\]

The $\ell$-step ahead forecasts, given $\mathbf{z}(t|t)$ are given by

\[
\begin{bmatrix}
\mathbf{x}_{1t}(\ell) \\
\mathbf{x}_{2t}(\ell)
\end{bmatrix} = \begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1
\end{bmatrix}^\ell \begin{bmatrix}
\mathbf{z}_1(t|t) \\
\mathbf{z}_2(t|t)
\end{bmatrix}
\]
\[\begin{bmatrix}
\phi_{11} x_1(t|t) + \phi_{12} \sum_{k=0}^{j-1} \phi_{11}^k x_2(t|t) \\
\end{bmatrix}
\]

and the corresponding MMSE matrices are given by

\[
\Sigma_x(t) = \sum_{j=0}^{\ell-1} \begin{bmatrix}
\phi_{11} & \phi_{12} \\
0 & 1 \\
\phi_{12} & 1 \\
\end{bmatrix}^j \begin{bmatrix}
\phi_{11} \\
0 \\
\phi_{12} \\
\end{bmatrix}^j
\]

To describe forecast regions, we take \(\phi_{11} = 0.5, \phi_{12} = 0.3\) and for \(\ell = 1, 2, 3\) we need

\[
W^0 = I = W^0(W')^0, \quad W = \begin{bmatrix}
0.5 & 0.3 \\
0 & 1 \\
\end{bmatrix}
\]

\[
W^2 = \begin{bmatrix}
0.25 & 0.45 \\
0 & 1 \\
\end{bmatrix}, \quad W^2(W')^2 = \begin{bmatrix}
0.265 & 0.45 \\
0.45 & 1 \\
\end{bmatrix}
\]

\[
W^3 = \begin{bmatrix}
0.125 & 0.525 \\
0 & 1 \\
\end{bmatrix}, \quad WW' = \begin{bmatrix}
0.34 & 0.3 \\
0.3 & 1 \\
\end{bmatrix}
\]

\[
\ell=1 \\
\begin{bmatrix}
x_{1t}(1) \\
x_{2t}(1) \\
\end{bmatrix} = \begin{bmatrix}
0.5 & 0.3 \\
0 & 1 \\
\end{bmatrix} \begin{bmatrix} 1 \\
1 \\
\end{bmatrix} = \begin{bmatrix} 0.8 \\
1 \\
\end{bmatrix}, \quad \Sigma_x(1) = I
\]

The 95\% confidence region is described by

\[(x - 0.8)^2 + (y - 1)^2 \leq 5.99\]

a circle centred \((0.8, 1)^t\) and radius 2.447.
$l=2$

\[
\begin{bmatrix}
  x_{1t}(2) \\
  x_{2t}(2)
\end{bmatrix} = \begin{bmatrix}
  0.25 & 0.45 \\
  0 & 1
\end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.7 \end{bmatrix}, \text{ centre.}
\]

\[
\Sigma_2(2) = \sum_{j=0}^{1} W^j(W')^j
\]

\[
= \begin{bmatrix} 1.34 & 0.3 \\ 0.3 & 2 \end{bmatrix}
\]

\[
v_1 = \begin{bmatrix} 0.361 \\ 0.933 \end{bmatrix}, \qquad \theta = 68.85^0
\]

eigenvalues: $\lambda^2 - 3.34\lambda + 2.59 = 0$

$\lambda_1 = 2.116, \quad \lambda_2 = 1.22$

length of major axis: $c \sqrt{\lambda_1} = 3.560$

length of minor axis: $c \sqrt{\lambda_2} = 2.703$

We see shift in centre towards the $x_{2t}$ axis along the line $x_{2t} = 1$, both major and minor axes expand but major axis expands quicker.

$l=3$

centre \[
\begin{bmatrix}
  x_{1t}(3) \\
  x_{2t}(3)
\end{bmatrix} = \begin{bmatrix}
  0.5 & 0.3 \\
  0 & 1
\end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.65 \end{bmatrix}
\]

\[
\Sigma_2(3) = \sum_{j=0}^{2} W^j(W')^j
\]

\[
= \begin{bmatrix} 1.605 & 0.75 \\
  0.75 & 3 \end{bmatrix}
\]

\[
v_1 = \begin{bmatrix} 0.399 \\ 0.75 \end{bmatrix}, \qquad \theta = 66.5^0
\]

eigenvalues: $\lambda^2 - 4.605\lambda + 4.2525 = 0$

$\lambda_1 = 3.3267, \quad \lambda_2 = 1.2783$

length of major axis: $c \sqrt{\lambda_1} = 4.464$

length of minor axis: $c \sqrt{\lambda_2} = 2.767$
The centre shifts to the left along $x_{2t} = 1$, the region increases on both axes.

**Example 5.4.3**

The next process is

$$
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix}
= \begin{bmatrix}
    1 & \phi_{12} \\
    0 & 1
\end{bmatrix}
\begin{bmatrix}
    x_{1,t-1} \\
    x_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}, \quad \phi_{12} \in \mathbb{R} \tag{5.4.4}
$$

The process (5.4.4) can be written

$$
x_{1t} - x_{1,t-1} = \phi_{12} x_{2,t-1} + \varepsilon_{1t} \tag{5.4.4a}
$$

$$
x_{2t} - x_{2,t-1} = \varepsilon_{2t} \tag{5.4.4b}
$$

so that $x_{2t} \sim I(1)$. The presence of $\phi_{12} x_{2,t-1}$ on the right hand side of (5.4.4a) implies that $x_{1t} - x_{1,t-1}$ is not a stationary process. In fact

$$
(x_{1t} - x_{1,t-1}) - (x_{1,t-1} - x_{1,t-2}) = \phi_{12} \varepsilon_{2,t-1} + (\varepsilon_{1t} - \varepsilon_{1,t-1}) \tag{5.4.4c}
$$

so that $x_{1t} \sim I(2)$. The process $\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$ is therefore not a cointegrated process. Expressions (5.4.4c) and (5.4.4b) show that $x_{1t}$ and $x_{2t}$ are an IMA(2, 1) process and a random walk respectively.

We choose $\phi_{12} = 1.5$ and we look at forecasts for $\ell = 1, 2, 3$.

Define

$$W^0 = I.$$

Now

$$W = \begin{bmatrix} 1 & 1.5 \\ 0 & 1 \end{bmatrix}, \quad WW' = \begin{bmatrix} 3.25 & 1.5 \\ 1.5 & 1 \end{bmatrix}$$
\[ W^2 = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \quad W^2(W^t)^2 = \begin{bmatrix} 10 & 3 \\ 3 & 1 \end{bmatrix} \]

\[ W^3 = \begin{bmatrix} 1 & 4.5 \\ 0 & 1 \end{bmatrix} \]

Set
\[ \mathbf{x}(t|t) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma_x(1) = 1 \]

\( \ell = 1 \)
\[ \begin{bmatrix} \mathbf{z}_1t(2) \\ \mathbf{z}_2t(2) \end{bmatrix} = W\mathbf{x}(t|t) = \begin{bmatrix} 2.5 \\ 1 \end{bmatrix}, \text{ centre.} \]

Ellipse becomes a circle centered at \[ \begin{bmatrix} 2.5 \\ 1 \end{bmatrix} \], radius \( \sqrt{5.99} \) for 95% forecast region.

\( \ell = 2 \)
\[ \begin{bmatrix} \mathbf{z}_1t(2) \\ \mathbf{z}_2t(2) \end{bmatrix} = W^2\mathbf{x}(t|t) = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \text{ centre.} \]

\[ \Sigma_{xx}(2) = \sum_{j=0}^{1} W^j(W^t)^j = \begin{bmatrix} 4.25 & 1.5 \\ 1.5 & 2 \end{bmatrix} \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \theta = 26.57^0 \]

Eigenvalues: \( \lambda^2 - 6.25\lambda + 6.25 = 0 \)
\( \lambda_1 = 5, \quad \lambda_2 = 1.25 \)

Length on major axis: \( c\frac{\lambda_1}{4} = 5.473 \)

Length on minor axis: \( c\frac{\lambda_2}{4} = 2.736 \)

Compared to the 1-step forecast region, the centre shifts towards the right along \( \mathbf{z}_{2t} = 1 \), and the lengths of major and minor axes expand at a high rate.
\[ \ell=3 \]

centre

\[
\begin{bmatrix}
W_0(x(t)) = [5.5] \\
W_1(x(t)) = [1]
\end{bmatrix}
\]

\[
\Sigma_x(3) = \begin{bmatrix}
14.25 & 4.5 \\
4.5 & 3
\end{bmatrix}, \quad v_1 = [0.944], \quad \theta = 19.3297^0
\]

eigenvalues:

\[\lambda_1^2 - 17.25\lambda + 22.5 = 0\]

\[\lambda_1 = 15.8285, \quad \lambda_2 = 1.4215\]

length of major axis: 9.737

length of minor axis: 2.918

The expansion is very significant along the major axis whereas relatively constant on minor axis. The centre shifts slowly along \(x_{2t} = 1\) to the right hand.

5.5 DISCUSSION

The circles described when \(\ell=1\) show that the expansion of forecasts for each process starts very slowly. All forecast regions expand but (5.4.4) is the quickest to expand. This is expected because two unit eigenvalues on \(\Phi\) are reflected while on (5.4.3) we chose \(|\phi_{11}| < 1\) so that only one unit eigenvalue is there. The slowest to expand is (5.4.2).

By checking the nature of expansion, the minor axis is (relatively) slower when comparing with major axis and this is the case with all the models. In fact, minor axis for (5.4.2) do not expand at all. This is due to lack of contribution of \(x_{1t}\). The angles \(\theta\) for (5.4.3) and (5.4.4) become smaller, this means that the ellipse for increasing \(\ell\) have major axis tending towards being parallel with the \(x_{1t}\) axis, whereas for (5.4.1) for \(\ell=2\) and \(\ell=3\), it stays around \(63^0\).
Lastly on (5.4.2), when \( a = 0 \) then we have ellipse along \( x_{2t} \) axis. The minor axis remain with fixed lengths.

Even though in all the models we focused on 95\% intervals, the values chosen for the \( \phi \)'s and \( a \) were not consistent, for example \( \phi_{12} = 1.5 \) in (5.4.4) and \( a = 0.5 \) (or \( \phi_{12} = 0.3 \)) in the other two models could be an expansion and a contraction (damp) respectively. The overall comparison has an unfair nature.
CHAPTER 6
COINTEGRATED BIVARIATE
STRUCTURAL TIME SERIES MODELS

The cointegrating properties of structural models are discussed using common factor model, or common structural components. Our discussion is limited to bivariate SUTSE structural models.

6.1 DYNAMIC FACTOR ANALYSIS

Factor analysis in bivariate time series is to set up a model set up with each of the two variables being a linear combination of one common factor, (Harvey (1989: 450) suggests up to two common factors for a bivariate process), plus a random disturbance term. For bivariate structural time series, the essence is to formulate them in terms of components which have distinct dynamic properties, that is components which can be interpreted. The SUTSE class are generalized by allowing them to have some of the components to be common. This topic has been discussed by many authors for stationary processes, but in this case we allow some of the components to be nonstationary. In fact the differences in the properties of components help to distinguish the components easily. For a bivariate case, any (bivariate) component where a common factor is introduced, one common factor will suffice.

6.1.1 Bivariate random walk plus noise with drift
Harvey (1989: 450)

As an introduction we look at a bivariate version of common trends model (8.5.1) in Harvey (1989: 450), where a bivariate random walk plus noise SUTSE model is being considered, and a drift \( \beta = (\beta_1 \beta_2)' \) also included.

A general bivariate model is
where the factors are \( \mu_{1t} \) and \( \mu_{2t} \).

From section 2.2.4 (Remarks) on bivariate random walk plus noise and bivariate local linear trend we deduce that the nonstationary part is the random walk because once it is deterministic there is no further nonstationarity. That is, (6.1.lb) is nonstationary, but

\[
\begin{bmatrix}
\Delta \mu_{1t} \\
\Delta \mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

is a white noise (stationary) process.

As a result (6.1.1a) is nonstationary with \( \Theta \mu_t \) as the nonstationary component, but

\[
\begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{12} & \theta_{22}
\end{bmatrix}
\begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
+ \begin{bmatrix}
\theta_{11} & \theta_{12} \\
\theta_{12} & \theta_{22}
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{1t} - \varepsilon_{1,t-1} \\
\varepsilon_{2t} - \varepsilon_{2,t-1}
\end{bmatrix}
\]

is stationary. That is, \( x_{1t} \sim I(1) \) and \( x_{2t} \sim I(1) \).

If cointegration exists between \( x_{1t} \) and \( x_{2t} \) it will be affected at (6.1.1a), and because \( \mu_t \) is the nonstationary component, the cointegrating matrices (or vectors) must remove the coefficient matrix \( \Theta \) of \( \mu \). That is if a cointegrating matrix \( A: 2 \times 2 \) exists (or \( \alpha: 2 \times 1 \)), it must satisfy the condition

\[
A \Theta = O_2 \quad (\alpha^T \theta = [0 \ 0])
\]

This is clear that \( \alpha \) are (multiples of) rows of \( A \), and that (in the latter
case \( \alpha' \theta = [0 \ 0] \) columns of \( \Theta \) are linearly dependent. Thus, rank \( \Theta = 1 \). Since not many \( \Theta \)'s have these properties, there are cases where no such \( A \) exists. That is, not all bivariate processes of the form (6.1.1) are cointegrated.

As an alternative, dynamic factor analysis may be applied on the nonstationary term, and in the next subsection we derive using factor analysis, the common level slope model.

### 6.1.2 Derivation of common level common slope model

Harvey (1989: 450), Lütkepohl (1991: 423). Help from Prof. Markham

In model (6.1.1a) we define \( \Theta \mu_t = \mu_t^* \) and apply dynamic factor analysis approach on the new level, that is (taking \( \mu_t \) as common for both \( \mu_{1t}^* \) and \( \mu_{2t}^* \))

\[
\begin{bmatrix}
\tau_{1t} \\
\tau_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t}^* \\
\mu_{2t}^*
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\mu_{1t}^* \\
\mu_{2t}^*
\end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mu_t + \begin{bmatrix} \eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

\[
\mu_t = \mu_{t-1} + \beta + v_t
\]

where the white noise processes \( \begin{bmatrix} \varepsilon_{1t} \\
\varepsilon_{2t}\end{bmatrix}, \begin{bmatrix} \eta_{1t} \\
\eta_{2t}\end{bmatrix} \) and \( v_t \) are uncorrelated with each other for all time periods, and each of them is itself uncorrelated at different time periods. Further, \( \eta_{1t} \) and \( \eta_{2t} \) are uncorrelated so that \( \Sigma_{\eta\eta} \) is diagonal, and \( \mu_t \) is uncorrelated with \( \begin{bmatrix} \eta_{1t} & \eta_{2t} \end{bmatrix}' \).

From previous sections, the random walk (with drift) \( \mu_t \) is nonstationary, leading us to conclude \( \mu_t^* \) is nonstationary and lastly that \( \tau_t \) is nonstationary.
Both $x_{1t}$ and $x_{2t}$ are nonstationary, and if we define

$$
\begin{bmatrix}
  w_{1t} \\
  w_{2t}
\end{bmatrix} = \begin{bmatrix}
  \eta_{1t} \\
  \eta_{2t}
\end{bmatrix} + \begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix}
$$

by substituting

$$
\begin{bmatrix}
  \mu_{1t}^* \\
  \mu_{2t}^*
\end{bmatrix} = \begin{bmatrix}
  1 \\
  0
\end{bmatrix} \mu_t + \frac{1}{\lambda} \begin{bmatrix}
  \eta_{1t} \\
  \eta_{2t}
\end{bmatrix}
$$

then

$$
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = \begin{bmatrix}
  1 \\
  0
\end{bmatrix} \mu_t + \begin{bmatrix}
  w_{1t} \\
  w_{2t}
\end{bmatrix}
$$

Then

$$
\begin{bmatrix}
  \Delta x_{1t} \\
  \Delta x_{2t}
\end{bmatrix} = \begin{bmatrix}
  \beta + \nu_t + \Delta w_{1t} \\
  \theta \beta + \theta \nu_t + \Delta w_{2t}
\end{bmatrix}
$$

is a stationary process (white noise of nonzero mean vector $[\beta \ \theta \beta]^\top$), and therefore $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$. Let $\alpha = [-\theta \ 1]^\top$, then

$$
\begin{bmatrix}
  -\theta & 1
\end{bmatrix} \begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = \begin{bmatrix}
  -\theta & 1
\end{bmatrix} \begin{bmatrix}
  w_{1t} \\
  w_{2t}
\end{bmatrix}
$$

is stationary so that $\alpha$ is a cointegrating vector.

Therefore, a required model with $\begin{bmatrix}
  -\theta & 1
\end{bmatrix} \begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix}$ stationary is

$$
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = \begin{bmatrix}
  1 \\
  0
\end{bmatrix} \mu_t + \begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix} \quad (6.1.2a)
$$

$$
\mu_t = \mu_{t-1} + \beta + \nu_t \quad (6.1.2b)
$$
where $\mu_t$ is the common level, $\beta$ the common slope.

From assumptions made about $\epsilon_t$ and $\eta_t$, it becomes true also that $v_t$ is uncorrelated with $\epsilon_t$.

### 6.1.3 Common slopes

We consider a structural time series model with trend and slope, but only the slope is allowed to have a common factor, that is

\begin{align}
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} &=
\begin{bmatrix}
    \mu_{1t} \\
    \mu_{2t}
\end{bmatrix}
+ \begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix} \quad (6.1.3a)
\end{align}

\begin{align}
\begin{bmatrix}
    \mu_{1t} \\
    \mu_{2t}
\end{bmatrix} &=
\begin{bmatrix}
    \mu_{1,t-1} \\
    \mu_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
    1 \\
    \theta \beta
\end{bmatrix} \begin{bmatrix}
    \beta_{t-1} \\
    \delta
\end{bmatrix}
+ \begin{bmatrix}
    0 \\
    \eta_{2t}
\end{bmatrix} + \begin{bmatrix}
    \eta_{1t}
\end{bmatrix} \quad (6.1.3b)
\end{align}

\begin{equation}
\beta_t = \beta_{t-1} + v_t \quad (6.1.3c)
\end{equation}

where the white noise processes are all uncorrelated with each other.

Because $\beta_t$ is nonstationary, this is a nonstationary process, now

\begin{align}
\begin{bmatrix}
    \Delta x_{1t} \\
    \Delta x_{2t}
\end{bmatrix} &=
\begin{bmatrix}
    1 \\
    \theta \beta
\end{bmatrix} \begin{bmatrix}
    \beta_{t-1} \\
    \delta
\end{bmatrix}
+ \begin{bmatrix}
    0 \\
    \eta_{2t} \quad \Delta \epsilon_{2t}
\end{bmatrix}
+ \begin{bmatrix}
    \eta_{1t} + \Delta \epsilon_{1t}
\end{bmatrix} \quad (6.1.3d)
\end{align}

is also nonstationary because of $\beta_t$. But

\begin{align}
\begin{bmatrix}
    \Delta^2 x_{1t} \\
    \Delta^2 x_{2t}
\end{bmatrix} &=
\begin{bmatrix}
    1 \\
    \theta \beta
\end{bmatrix} \begin{bmatrix}
    v_{t-1} \\
    \Delta \eta_{2t} \quad \Delta^2 \epsilon_{2t}
\end{bmatrix}
+ \begin{bmatrix}
    \Delta \eta_{1t} + \Delta^2 \epsilon_{1t}
\end{bmatrix} \quad (6.1.3e)
\end{align}

is stationary (white noise). That is $x_{1t} \sim I(2), \quad x_{2t} \sim I(2)$. 

Substituting (6.1.3b) in (6.1.3a) and multiply by \([\theta \beta \ -1]\) we obtain

\[
\theta \beta x_{1t} - x_{2t} = \theta \beta \mu_{1,t-1} - \mu_{2,t-1} - \delta + \theta \beta \eta_{1t} - \eta_{2t} + \theta \beta \epsilon_{1t} - \epsilon_{2t}
\]

which is nonstationary due to the trends \(\mu_{1,t-1}\) and \(\mu_{2,t-1}\).

But

\[
\Delta(\theta \beta x_{1t} - x_{2t}) = \theta \beta \beta_{1-2} + \theta \beta \eta_{1,t-1} - \theta \beta \eta_{2,t-1} + \theta \beta \eta_{1t} - \theta \beta \Delta \eta_{1t} - \Delta \eta_{2t} - \theta \beta \Delta \epsilon_{1t} - \Delta \epsilon_{2t}
\]

is stationary.

Therefore

\[
\begin{bmatrix}
\theta \beta \\
-1
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} \sim I(1)
\]

and therefore

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} \sim CI(2,1)
\]

and the cointegrating vectors are multiples of \([\theta \beta \ -1]\).

### 6.2 FORECASTS

We want to investigate the behavior of forecasts of the common level common slope model (6.1.2) and of the common slope model of (6.1.3). We use the SSR (2.3.1) and for point forecasts and MMSE we require (2.3.15) and (2.3.17) respectively.
6.2.1 Forecasts for common level, common slope model

The model is (6.1.2) given by

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    1 & \mu_t \\
    0 & \theta
\end{bmatrix} 
\begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

\[
\mu_t = \mu_{t-1} + \beta + v_t
\]

Let

\[
\begin{bmatrix}
    z_{1t} \\
    z_{2t}
\end{bmatrix} = \begin{bmatrix}
    \mu_t \\
    \beta
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    \mu_{t-1} + \beta + v_t \\
    \beta
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    1 & 1 & \mu_{t-1} \\
    0 & 1 & \beta
\end{bmatrix} + \begin{bmatrix}
    1 & v_t \\
    0 & 0
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = \begin{bmatrix}
    1 & \mu_t \\
    0 & \theta
\end{bmatrix} 
\begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
    1 & 0 & \mu_t \\
    0 & 0 & \beta
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{1t} \\
    \varepsilon_{2t}
\end{bmatrix}
\]

Let \( \varepsilon_t^{(1)} \) and \( \varepsilon_t^{(2)} \) be white noise processes with zero mean vectors and covariance matrix \( I_2 \) each. We note from factor analysis that \( \varepsilon_{1t} \) and \( \varepsilon_{2t} \) are uncorrelated then \( \Sigma_{\varepsilon\varepsilon} \) is diagonal. We write the SSR as

\[
\begin{bmatrix}
    z_{1t} \\
    z_{2t}
\end{bmatrix} = \begin{bmatrix}
    1 & 1 & z_{1,t-1} \\
    0 & 1 & z_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
    \sigma_v & 0 & 0 \\
    0 & 0 & 0
\end{bmatrix} \begin{bmatrix}
    \varepsilon_t^{(1)} \\
    \varepsilon_t^{(2)}
\end{bmatrix}
\]
We define $W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. We define $W^0 = I_2$.

$W^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, $W^3 = W^2W = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$.

We proceed by induction, as $W^j$ for $j = 1, 2, 3$ have been found as

$W^j = \begin{bmatrix} 1 & j \\ 0 & 1 \end{bmatrix}$

we assume that the induction step below is also true, that is

$W^{k-1} = \begin{bmatrix} 1 & k-1 \\ 0 & 1 \end{bmatrix}$

Based on the induction step,

$W^k = W^{k-1}W$

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

and therefore (using also $W^0 = I$), $W^k$ is true as given above for $k = 0, 1$, and so on.

$HW^k = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & k \\ 0 & 0 \end{bmatrix}$

$= \begin{bmatrix} 1 & k \\ 0 & 0 \end{bmatrix}$, $k = 0, 1, 2, \ldots$

$GG^j = \begin{bmatrix} \sigma_{11,\varepsilon} & 0 \\ 0 & \sigma_{22,\varepsilon} \end{bmatrix}$
\[ HW^k_B = \begin{bmatrix} 1 & k \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_x & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ = \begin{bmatrix} \sigma_x & 0 & 0 & 0 \\ 0 & \sigma_x & 0 & 0 \end{bmatrix} \]

\[ HW^k_B(HW^k_B)^t = \sigma_x^2 \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 \end{bmatrix}, \quad k = 0, 1, \ldots \]

Suppose that 
\[ z(t|t) = [\mu_1(t|t)] \quad \text{and} \quad \Sigma_{zz}(t|t) = \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix} \]

are given, then the \( \ell \)-step ahead forecasts are

\[
\begin{bmatrix} x_{1t}(\ell) \\ x_{2t}(\ell) \end{bmatrix} = HW^\ell z(t|t)
\]

\[ = \begin{bmatrix} 1 & \ell \\ 0 & \ell \end{bmatrix} \begin{bmatrix} \mu_1(t|t) \\ \beta(t|t) \end{bmatrix}
\]

\[ = \begin{bmatrix} \mu_1(t|t) \\ \theta \mu_1(t|t) \end{bmatrix} + \ell \begin{bmatrix} \beta(t|t) \\ \theta \beta(t|t) \end{bmatrix}, \quad \ell = 1, 2, \ldots \]

and the corresponding MMSE's are given by:

\[
\Sigma_{xx}(\ell) = HW^\ell \Sigma_{zz}(t|t)(HW^\ell)^t + \sum_{j=0}^{\ell-1} HW^j B(HW^j B)^t + GG^t
\]

\[ \ell = 1 \]

\[
\Sigma_{xx}(1) = HW \Sigma_{zz}(t|t)(HW)^t + HBB^t H + GG^t
\]
\[
= (\sigma_{11}(t|t) + 2\sigma_{12}(t|t) + \sigma_{22}(t|t)) [1 \quad 0] + \sigma_{\nu}^2 [1 \quad 0] \\
+ [\sigma_{11,\varepsilon} \quad 0] \\
\begin{bmatrix} \sigma_{22,\varepsilon} \end{bmatrix}
\]

\[\ell \in \mathbb{N}, \ell \geq 2\]
\[
\Sigma_{xx}(\ell) = \Pi W^\ell \Sigma_{xx}(t|t)(HW^\ell)' + \sum_{j=0}^{\ell-1} \Pi HW^j B(HW^j B)' + GG'
\]
\[
= (\sigma_{11}(t|t) + 2\ell \sigma_{12}(t|t) + \ell^2 \sigma_{22}(t|t)) [1 \quad 0] + \sum_{j=0}^{\ell-1} \Sigma \sigma_{\nu}^2 [1 \quad 0] \\
+ [\sigma_{11,\varepsilon} \quad 0] \\
\begin{bmatrix} \sigma_{22,\varepsilon} \end{bmatrix}
\]

Therefore combining with \(\ell=1\),
\[
\Sigma_{xx}(\ell) = (\sigma_{11}(t|t) + 2\ell \sigma_{12}(t|t) + \ell^2 \sigma_{22}(t|t) + \ell \sigma_{\nu}^2) [1 \quad 0] \\
+ [\sigma_{11,\varepsilon} \quad 0] \\
\begin{bmatrix} \sigma_{22,\varepsilon} \end{bmatrix}, \ell \in \mathbb{N}, \ell \geq 1
\]

We defined the forecast regions in Chapter 5, and now we need to evaluate parameters first. We assume that the errors \(\mu(t) - \mu(t|t)\) and \(\beta - \beta(t|t)\) are uncorrelated. We further assume that all variances are equal to one, and we set \(\theta = 0.5, \beta(t|t) = 1.5\). Let \(z_1(t|t) = 1\) be given, then
\[
\begin{bmatrix}
    z_{1t}(\ell) \\
    z_{2t}(\ell)
\end{bmatrix} = 
\begin{bmatrix}
    1 + 1.5\ell \\
    1.5 + 0.75\ell
\end{bmatrix}, \quad \ell = 1, 2, ...
\]

and

\[
\Sigma_{zz}(\ell) = (1+\ell+\ell^2) \begin{bmatrix}
    1 & 0.5 \\
    0.5 & 0.25
\end{bmatrix} + 
\begin{bmatrix}
    1 & 0 \\
    0 & 1
\end{bmatrix}, \quad \ell = 1, 2, ...
\]

For the following forecast regions we will use Johnson & Wichern (1992: 132) method of plotting ellipses. 90% regions implies \( c^2 = \chi^2_{2}(0.10) = 4.61.\)

\(\ell=1\)

\[
\begin{bmatrix}
    x_{1t}(1) \\
    x_{2t}(1)
\end{bmatrix} = 
\begin{bmatrix}
    2.5 \\
    2.25
\end{bmatrix}, \quad \Sigma_{zz}(1) = 
\begin{bmatrix}
    4.00 & 1.50 \\
    1.10 & 1.75
\end{bmatrix} \quad v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix}
    2 \\
    1
\end{bmatrix}, \quad \theta = 26.565^0
\]

eigenvalues: \( \lambda^2 - 5.75\lambda + 4.75 = 0 \)
\( \lambda_1 = 4.75, \lambda_2 = 1 \)

length of major axis: \( c\sqrt{\lambda_1} = 4.679 \)

length of minor axis: \( c\sqrt{\lambda_2} = 2.147 \)

\(\ell=2\)

\[
\begin{bmatrix}
    x_{1t}(2) \\
    x_{2t}(2)
\end{bmatrix} = 
\begin{bmatrix}
    4 \\
    3
\end{bmatrix}, \quad \Sigma_{zz}(2) = 
\begin{bmatrix}
    8.00 & 3.50 \\
    3.50 & 1.75
\end{bmatrix} \quad v_1 = \begin{bmatrix}
    0.913 \\
    0.409
\end{bmatrix}, \quad \theta = 24.13^0
\]

eigenvalues: \( \lambda_1 = 9.567, \lambda_2 = 0.183 \)

\( c\sqrt{\lambda_1} = 6.641, \quad c\sqrt{\lambda_2} = 0.918 \)

The centre shifts both ways; that is up the \( x_{2t} \) and right-ways along the \( x_{1t} \). The minor axis has fixed length so far, and the major axis increases.
\[ L = 5 \]

\[
\begin{bmatrix}
\mathbf{x}_{1t}(5) \\
\mathbf{x}_{2t}(5)
\end{bmatrix} =
\begin{bmatrix}
8.5 \\
5.25
\end{bmatrix},
\begin{bmatrix}
\Sigma_{xx}(5)
\end{bmatrix} =
\begin{bmatrix}
32.00 & 15.50 \\
15.50 & 7.75
\end{bmatrix},
\begin{bmatrix}
\mathbf{v}_1 \\
\theta
\end{bmatrix} =
\begin{bmatrix}
0.899 \\
25.98^0
\end{bmatrix}
\]

\[
\lambda_1 = 39.554, \quad \lambda_2 = 0.196
\]

\[ c \sqrt{\lambda_1} = 13.503, \quad c \sqrt{\lambda_2} = 0.950 \]

We realize from previous \( L \)'s that the \( \theta \) increases, therefore as \( L \) increases the ellipse rotates anticlockwise, hence in the long run it approximates ellipse parallel to \( x_{2t} \) axis. The centre has shifted, still both \( x_{1t} \) and \( x_{2t} \) increasing. The major axis increase in length for increasing \( L \), while the minor axis has fixed lengths.

### 6.2.2 Forecasts for the common slope model

The model is (6.1.3) which is given by

\[
\begin{bmatrix}
\mathbf{x}_{1t} \\
\mathbf{x}_{2t}
\end{bmatrix} =
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} =
\begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
1 \\
0
\end{bmatrix} \beta_{t-1} +
\begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

\[
\beta_t = \beta_{t-1} + \mathbf{w}_t
\]

where factor analysis has \( \Sigma_{\mathbf{\eta} \mathbf{\eta}} = \text{diag}(\sigma_{11}, \sigma_{22}, \eta) \), and by nature of structural models, \( \mathbf{w}_t, \eta_t \) and \( \varepsilon_t \) are uncorrelated.

We derive the SSR for this process.

Define
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \begin{bmatrix}
  \mu_{1t} \\
  \mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
  \mu_{1,t-1} + [1] \beta_{t-1} + [0] \delta + [\eta_{1t}] \\
  \mu_{2,t-1} + [\theta \beta] + [1] [\eta_{2t}]
\end{bmatrix}
\]

Define
\[
\begin{bmatrix}
  z_{3t} \\
  z_{4t}
\end{bmatrix} = \begin{bmatrix}
  \beta_t \\
  \delta
\end{bmatrix}
\]

then
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix} = \begin{bmatrix}
  z_{1,t-1} + [1] 0 [z_{3,t-1}] + [\eta_{1t}] \\
  z_{2,t-1} + [\theta \beta] 1 [z_{4,t-1}] + [\eta_{2t}]
\end{bmatrix}
\]

Recursion for
\[
\begin{bmatrix}
  z_{3t} \\
  z_{4t}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  z_{3t} \\
  z_{4t}
\end{bmatrix} = \begin{bmatrix}
  \beta_{t-1} + w_t \\
  \delta
\end{bmatrix}
= \begin{bmatrix}
  \beta_{t-1} \\
  \delta
\end{bmatrix} + \begin{bmatrix}
  w_t \\
  0
\end{bmatrix}
\]

Let \( v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \) be white noise with \( \Sigma_{vv} = I_5 \), then
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t} \\
  \cdots \\
  z_{3t} \\
  z_{4t}
\end{bmatrix} = \begin{bmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & \theta \beta & 0 \\
  \cdots & \cdots & \cdots & \cdots \\
  0 & 0 & 0 & 1 \\
  0 & 0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
  z_{1,t-1} \\
  z_{2,t-1} \\
  \cdots \\
  z_{3,t-1} \\
  z_{4,t-1}
\end{bmatrix} + \begin{bmatrix}
  0 & 0 & \sqrt{\sigma_{11,\eta}} & 0 & 0 \\
  0 & 0 & 0 & \sqrt{\sigma_{22,\eta}} & 0 \\
  \cdots & \cdots & \cdots & \cdots & \cdots \\
  0 & 0 & 0 & 0 & \sqrt{\sigma_{ww}}
\end{bmatrix} v_t
\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} = 
\begin{bmatrix}
    1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0
\end{bmatrix} 
\begin{bmatrix}
    z_{1t} \\
    z_{2t} \\
    z_{3t} \\
    z_{4t}
\end{bmatrix} + \begin{bmatrix}
    \Sigma_{E}^{\frac{1}{2}} \\
    O_2
\end{bmatrix} \begin{bmatrix}
    I_2 \\
    I_2
\end{bmatrix} v_t
\]

Define \( W^0 = I_4 \)

\[
W = \begin{bmatrix}
    1 & 0 & 1 & 0 \\
    0 & 1 & \theta \beta & 1 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}, \quad
W^2 = \begin{bmatrix}
    1 & 0 & 2 & 0 \\
    0 & 1 & 2\theta \beta & 2 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
\]

\[
W^3 = \begin{bmatrix}
    1 & 0 & 3 & 0 \\
    0 & 1 & 3\theta \beta & 3 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
\]

By using induction we can show that

\[
W^k = \begin{bmatrix}
    1 & 0 & k & 0 \\
    0 & 1 & k\theta \beta & k \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}, \quad k = 0, 1, 2, ...
\]

\[
HW^k = \begin{bmatrix}
    1 & 0 & k & 0 \\
    0 & 1 & k\theta \beta & k
\end{bmatrix}, \quad k = 0, 1, 2, ...
\]

\[
HW^kB = \begin{bmatrix}
    0 & 0 & \sqrt{\sigma_{11,\eta}} & 0 & k\sqrt{\sigma_{ww}} \\
    0 & 0 & 0 & \sqrt{\sigma_{22,\eta}} & k\theta \beta \sqrt{\sigma_{ww}}
\end{bmatrix}
\]

\[
HW^kB(HW^kB)^t = \begin{bmatrix}
    \sigma_{11,\eta} + k^2 \sigma_{ww} & k^2 \theta \beta \sigma_{ww} \\
    k^2 \theta \beta \sigma_{ww} & \sigma_{22,\eta} + k^2 \theta^2 \beta \sigma_{ww}
\end{bmatrix}
\]

Let \( z(t|t) \) and \( \Sigma_{zz}(t|t) \) be given, then
\[
\begin{align*}
[x_{1t}] &= HW^t \nu(t \mid t) \\
[x_{2t}] &= \begin{bmatrix} 1 & \ell & 0 \\ 0 & 1 & \ell \theta \beta \end{bmatrix} \begin{bmatrix} z_1(t \mid t) \\ z_2(t \mid t) \\ z_3(t \mid t) \\ z_4(t \mid t) \end{bmatrix} \\
&= \begin{bmatrix} x_1(t \mid t) + \ell \beta(t \mid t) + \ell \delta(t \mid t) \\ x_2(t \mid t) + \theta \beta \delta(t \mid t) \end{bmatrix}
\end{align*}
\]

and by using the partitioning

\[
\Sigma_{zz}(t \mid t) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{pmatrix}
\]

then

\[
\Sigma_z(t) = HW^t \Sigma_{zz}(t \mid t) (HW^t)^t + \sum_{j=0}^{\ell-1} HW^j B (HW^j B)^t + GG^t, \quad \ell \in \mathbb{N}, \quad \ell \geq 1
\]

where \( GG^t = \Sigma_{\varepsilon \varepsilon} \)

\[
\begin{align*}
\sum_{j=0}^{\ell-1} HW^j B (HW^j B)^t &= \ell \sigma_{11, \eta} + \sigma_{ww} \sum_{j=1}^{\ell-1} J^2 + \theta \beta \sigma_{ww} \sum_{j=1}^{\ell-1} J^2 \\
&\quad + \theta \beta \sigma_{ww} \sum_{j=1}^{\ell-1} J^2 + \ell \sigma_{22, \eta} + \theta \beta \sigma_{ww} \sum_{j=1}^{\ell-1} J^2 \\
&= \frac{\sigma_{ww}}{6} (\ell-1) \ell (2\ell-1) \begin{bmatrix} 1 & \theta \beta \\ \theta \beta & \theta^2 \beta \end{bmatrix} + \ell \begin{bmatrix} \sigma_{11, \eta} & 0 \\ 0 & \sigma_{22, \eta} \end{bmatrix}
\end{align*}
\]
For all random components we assume variance 1, and lack of pairwise correlation. Also \( \Sigma_{\varepsilon\varepsilon} = I_2 \). For the forthcoming calculations we set \( \theta_\beta = 0.6 \), \( \delta(t | t) = 1.4 \) and suppose \( z_1(t | t) = z_2(t | t) = z_3(t | t) = 1 \). Now we are given

\[
\begin{bmatrix}
1 \\
1 \\
1 \\
1.4
\end{bmatrix}
\]

Then the formulae become:

\[
\begin{bmatrix}
x_1(t) \\
x_2(t)
\end{bmatrix} = 
\begin{bmatrix}
1 + \ell \\
1 + 0.6\ell + 1.4\ell
\end{bmatrix} = 
\begin{bmatrix}
1 + \ell \\
1 + 2\ell
\end{bmatrix}, \text{ centres and}
\]

\[
\Sigma_{xx}(\ell) = 
\begin{bmatrix}
1 + \ell^2 & 0.6\ell^2 \\
0.6\ell^2 & 1 + 1.36\ell^2
\end{bmatrix} + (\ell - 1)\ell(2\ell - 1)
\begin{bmatrix}
0.16 & 0.10 \\
0.10 & 0.06
\end{bmatrix}
\]

\[
+ \ell
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

\[
\ell = 1 \quad (c = \sqrt{4.61} \text{ for } 90\%)
\]

\[
\begin{bmatrix}
x_1(t) \mid (\ell) \\
x_2(t) \mid (\ell)
\end{bmatrix} = 
\begin{bmatrix}
2 \\
3
\end{bmatrix}, \quad \Sigma_{xx} \mid (\ell) = 
\begin{bmatrix}
4.00 & 0.60 \\
0.60 & 4.36
\end{bmatrix}, \quad v_1 = [0.597], \quad \theta = 53.34^0
\]

\[
\lambda_1 = 4.807, \quad \lambda_2 = 3.554
\]

\[
c \sqrt{\lambda_1} = 4.707, \quad c \sqrt{\lambda_2} = 4.048
\]
This forecast region is already wider (both major and minor axis) than the counterpart \( x_t(1) \) of common level common slope. The angle \( \theta \) is also closer to \( 90^0 \) than its counterpart.

\[ t = 2 \]

\[
\begin{bmatrix}
 x_{1t} (2) \\
 x_{2t} (2)
\end{bmatrix} = \begin{bmatrix}
 3 \\
 5
\end{bmatrix}, \quad \Sigma_{xx} (2) = \begin{bmatrix}
 9.00 & 3.00 \\
 3.00 & 9.80
\end{bmatrix}, \quad v_1 = \begin{bmatrix}
 0.659
\end{bmatrix}, \quad \theta = 48.77^0
\]

\[ \lambda_1 = 12.427, \quad \lambda_2 = 6.374 \]

\[ c_{\lambda_1} = 7.569, \quad c_{\lambda_2} = 5.421 \]

The \( \theta \) becomes smaller to suggest that the ellipse rotates clockwise, unlike the anticlockwise rotation displayed by common level common slope. The centre is shifted up and to the right. As expected from the forms of MMSE's, the expansion is quicker than in the previous model. The major axis is even far quicker.

\[ t = 5 \]

\[
\begin{bmatrix}
 x_{1t} (5) \\
 x_{2t} (5)
\end{bmatrix} = \begin{bmatrix}
 6 \\
 11
\end{bmatrix}, \quad \Sigma_{xx} (5) = \begin{bmatrix}
 62.0 & 33.0 \\
 33.0 & 51.8
\end{bmatrix}, \quad v_1 = \begin{bmatrix}
 0.759
\end{bmatrix}, \quad \theta = 40.62^0
\]

\[ \lambda_1 = 90.292, \quad \lambda_2 = 23.508 \]

\[ c_{\lambda_1} = 20.402, \quad c_{\lambda_2} = 10.410 \]

The regions expand very quickly, but more along major axis. The minor axis expands slowly. The centre is consistent in its movement. The shift is up and right directions.
6.3 DISCUSSION

6.3.1 Interpretation of variables
The variables \( x_{1t} \) and \( x_{2t} \) are given in terms of parameters \( \mu_{1t} \) and these, depending on the nature of the structure (behavior), are given by factors such as \( \beta, \beta_t, \delta \) and so on, which as Johnson & Wichern (1992: 443) would suggest, "specify phenomena in terms of their presumed cause-and-effect variables". In other words, the parameters have a 'direct interpretation'. Of course these parameters are analysed and interpreted to give meaning to the variables \( x_{1t} \). The interpretation of a (bivariate) time series therefore depends entirely on its parameters. To interpret variables is to describe their behavior independently, and relative to each other. This is where (causality and) cointegration come in. Time series analysis is having forecasting as one of its objectives, we need interpretation of variables to be able to anticipate the future, that is, to prepare forecasts.

A cointegrated (structural time series) process is described by a common factor model, and as this common factor affects both series in the same way, we need to interpret it only once. As the series never drift 'too' far apart, one way interpretation of the common effect (factor) will explain how the series are kept always close. Through a cointegrating vector we are able to regress one variable on the other, and once one of the series has been interpreted, the other one may be considered a response variable. Unfortunately this latter approach, as it has thrown away the other factor(s), it does not necessarily reveal the long-term relation between the variables.

The cointegrated bivariate model has nonstationary components. An interesting property is that no matter how nonstationary they are, their nonstationarity is very limited when the two series are considered relative to each other.

The components of a cointegrating vector in a bivariate set up (with one common factor) are both nonzero, otherwise one variable vanishes when \( [0 \ 0] \) or \( [\theta \ 0] \) is premultiplying the process and no relation between variables can be reflected. Also, if we take
where $\mu_t$ is the common factor which is nonstationary, the cointegrating vector when used, removes the common factor by removing its loading or coefficient vector. The nonstationarity of $\mu_t$ is contained within fixed bounds (or removed), where

$$\alpha' [\theta_1, \theta_2] = 0$$

for every cointegrating vector $\alpha$. Conversely, because the cointegrating vector is always aimed at the common factor to reduce/remove nonstationarity, it is the common factor which contains most/all of the nonstationarity. Also, the cointegrating vector is always orthogonal to the coefficient vector of the common factor.

### 6.3.2 Point forecasts and forecast regions

As read off from the equations, the common slope diverges quicker than the common level common slope in point forecasts because of excess term $\ell \delta$. The forecast region for this model expands much quicker than for the common level common slope. Because in the common level common slope there are more components which are used to contain the nonstationarity, it expands slowly, whereas in the common slope, only the slope is expected to contain it. If there is no component to contain this behavior, the expansion would be still even quicker. The expansion does not mean any problem with the model, but only that the nonstationarity behavior can be monitored for certain horizons $\ell$, but in the long run we will not have an idea as to where our forecast can be located.

The nonstationarity, even for cointegrated models, is always reflected either in point forecasts or MMSE’s, or both. As we saw in forecast regions, some axis would increase slowly or stay fixed, and that is the effect of cointegration. In the wandering of these process nonetheless, if there is cointegration, the two variables wander together. The region which expands only one axis, displays a ‘strip’ within which both variables will always be found together.
Harvey (1989: 468-9)

A bivariate AR(p) model is given by

\[
\sum_{j=0}^{p} \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix} L^j \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\]

where

\[
\begin{bmatrix} \phi_{11,0} & \phi_{12,0} \\ \phi_{12,0} & \phi_{22,0} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]

A serious problem with this model is that it has a very large number of parameters. When the samples are very large (which does not occur very often), Harvey (1989: 468) states that 'the estimates are fairly efficient'. One possibility again, is that the roots of the determinant of the AR matrix, \(|\Phi(z)| = 0\), may not all be larger than one in absolute value. In this case the stationary condition is not satisfied and therefore the series is nonstationary. Harvey (1989: 469) discourages fitting of autoregressions to nonstationary series (especially) in small samples as this "can lead to serious problems, detrending can lead to considerable distortion, and that differencing does not in general provide a satisfactory way of fitting bivariate AR's to nonstationary series".

Suppose that the bivariate ARMA process
has no root on the unit circle, then the AR matrix is nonsingular and its inverse is written

\[
\begin{bmatrix}
\phi_{11}(L) & \phi_{12}(L) \\
\phi_{21}(L) & \phi_{22}(L)
\end{bmatrix}^{-1} = \frac{1}{\Phi(L)} \begin{bmatrix}
\phi_{22}(L) & -\phi_{12}(L) \\
-\phi_{21}(L) & \phi_{11}(L)
\end{bmatrix}
\]

Premultiplying the ARMA process by this inverse and manipulating of the equation leads to the final form

\[
|\Phi(L)| \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \text{adj} \Phi(L) \Theta(L) \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}
\]

where

\[
\text{adj} \Phi(L) = \begin{bmatrix}
\phi_{22}(L) & -\phi_{12}(L) \\
-\phi_{21}(L) & \phi_{11}(L)
\end{bmatrix}
\]

By writing \( \phi^*(L) = |\Phi(L)| \) then each series has the same autoregressive model and the right hand side of (7.0.1) consists of MA polynomials of order \( p+q \). Thus, each of the time series components in model (7.0.1) can be written as ARMA process with the same autoregressive component \( \phi^*(L) \):

\[
\phi^*(L) z_{it} = \theta^*_i(L) \epsilon_{it}, \quad i = 1, 2
\]

To say 'detrrending of a time series' means to regress it on \( t \), that is \( z_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_k t^k + \epsilon_t \), and for a bivariate time series the same form is correct but the \( \alpha \)'s are \( 2 \times 1 \) vectors.

In the next example we discuss the predator-prey relationship between two animals, mink (predator) and muskrat (prey). The study of this relationship is based on the hypothesis that mink is a predator of muskrat, and a guideline used is that if the predator-prey relationship does exist, then more mink species should be followed by few muskrats in the next time period (as
they will be reduced by minks having fed themselves), and few muskrats be followed by few minks (as minks did not have enough food they decrease). As minks decrease muskrats increase which means few minks are followed by many muskrats. As minks have enough food they increase, and the cycle is repeated.

We look at the statistical investigation of this relationship as studied by Chan & Wallis (1978), Harvey (1981b, 1989), Reinsel (1993) and others. The data used for this study are the Hudson's Bay Company’s annual fur sales from 1850 till 1911.

Example 7.1 (Nonstationary series) (Mink–muskrat example)
The example was done by Jenkins (1975), Chan & Wallis (1978), Harvey (1981b), Jenkins & Alavi (1981), and Cooper & Wood (1982) using the bivariate ARIMA model. A problem that exists is that mink observations are stationary while muskrat ones are I(1). To difference both series once is to overdifference the mink data which do not require any differencing, while to difference muskrat alone would distort the actual relationships. Both the original data and their logarithms were analysed (at different times by different analysts). Plots for both the original data and of their logarithms are presented.

(See Figure 7.1 on graph page 154a for plot of data.)

(See Figure 7.2 on graph page 154b for plot of logarithms.)
The original data as seen from Figure 7.1 are not conducive to comparison, graphs displayed are in such a way that when muskrat displays a clear picture, the mink is relatively hidden whereas the plot (Figure 7.2) on logarithms, even when the numbers are not equal, the graphs of both log-mink and log-muskrat are both readable and can be interpreted.

Chan & Wallis analysed the logarithm data, they detrended them by regressing each series on \( t \) and \( t^2 \). A number of competing ARMA models were fitted to seek the (most) suitable one. The one they settled for, which is the only one which can be put in the form (7.0.1), is the bivariate AR process:
Figure 7.1 Mink and muskrat furs sold by Hudson's Bay Company.
Figure 7.2 Logarithms of annual sales of mink and muskrat furs (gold) by Hudson's Bay Company.
\[
\begin{bmatrix}
1 - 0.79L & 0.69L \\
-0.29L & 1 - 0.51L
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

and when put in (final) form (7.0.1) we have

\[
\begin{bmatrix}
1 - 1.30L^2 + (0.603L^2) \\
-0.29L & 1 - 0.791L
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
1 - 0.51L & -0.69L \\
0.29L & 1 - 0.791L
\end{bmatrix}
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

The AR polynomial (or determinantal polynomial)

\[|\Phi(L)| = 1 - 1.30L^2 + 0.603L^2\]

has complex roots, and as this is for each series we recall that in univariate ARMA processes, as these complex roots are square roots of functions of the form \(e^{ix}\), we have the mixed sinusoidal-autoregressive model, and we may fit models of the form

\[y_t = \alpha \sin 2\pi f(t - \phi) + \beta y_{t-1} + \eta_t\]

which is designed "to capture both the regularity in the period of the oscillation in animal populations and the irregularity in their amplitude" as Chan & Wallis would suggest. (It could still be cosine in the place of sine above.) Harvey (1981b: 188-9) concludes that "this yields stochastic cycles similar to those observed in the data". The AR(1) process reproduces the lead-lag structure which signals the predator-prey relationship through its off-diagonal elements. These off-diagonal elements (through the positive and negative signs of coefficients 0.69 and -0.29) imply that an increase in muskrat in year \(t\) is followed by a decrease in muskrat in year \(t+1\), and this is then followed by a decrease in muskrat in year \(t+2\), which is followed by a decrease in mink in year \(t+3\), and so on.

Differencing both series could not be a good option as the two series have different orders of integration. Mink does not require any differencing to achieve stationarity as it is already stationary, while muskrat requires differencing once. As we seek a relationship between them we should consider it necessary to have a neutral common starting point for both series. We
therefore avoid differencing muskrat alone, and also not to overdifference mink, taking logarithms for both and detrending on same regressors $t$ and $t^2$ became a better option. As suggested, Figure 7.1 shows that comparison of original data poses a threat to accuracy as the scores of muskrat are very large in comparison to mink scores, and logarithms bring them closer for convenience in the handling of these sets.

Harvey (1981b, 1989) concludes that an autoregressive model does not provide a reasonable approximation.

**Example 7.2 (Cointegrated series)**

(The following model comes from Harvey (1989: 469 (8.8.4))

Consider the series

$$
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix}
= 
\begin{bmatrix}
  x_{1,t-1} + \varepsilon_{1t} \\
  \beta x_{1t} + \varepsilon_{2t}
\end{bmatrix}
$$

$x_{1t}$ is a (nonstationary) random walk so that $x_{1t} \sim I(1)$ because

$\Delta x_{1t} = \varepsilon_{1t}$ is stationary,

as a linear function of $x_{1t}$, $x_{2t}$ is also nonstationary, but

$\Delta x_{2t} = \beta \varepsilon_{1t} + \Delta \varepsilon_{2t}$

which is stationary white noise and $x_{2t} \sim I(1)$.

That is, $x_{1t} \sim I(1), x_{2t} \sim I(1)$.

By writing

$$
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix}
= 
\begin{bmatrix}
  1 & 0 \\
  \beta & 1
\end{bmatrix}
\begin{bmatrix}
  x_{1,t-1} + \varepsilon_{1t} \\
  \beta x_{1t} + \varepsilon_{2t}
\end{bmatrix}
$$

$$
= 
\begin{bmatrix}
  1 \\
  \beta
\end{bmatrix}
\begin{bmatrix}
  x_{1,t-1} \\
  \varepsilon_{1t}
\end{bmatrix}
+ 
\begin{bmatrix}
  1 \\
  \beta
\end{bmatrix}
\begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix}
$$
Let \( \alpha = \begin{bmatrix} \beta \\ -1 \end{bmatrix} \), then

\[
\alpha^t \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = [0 \ -1] \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = -\varepsilon_{2t}
\]

is a stationary white noise process, (or \( I(0) \)).

Therefore

\[
\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \sim CI(1,1)
\]

Let \( y_{1t} = \Delta z_{1t}, \ i = 1, 2 \) be stationary processes formed from the \( z_{1t} \), then the resulting ARMA(1,1) is

\[
\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \beta \Delta x_{1t} + \beta \Delta \varepsilon_{1t} + \Delta \varepsilon_{2t} \end{bmatrix}
\]

That is,

\[
\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{1,t-1} \\ \Delta x_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \beta \Delta & \Delta \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}
\]

The MA polynomial has determinant

\[
\begin{vmatrix} 1 & 0 \\ \beta(1-L) & 1-L \end{vmatrix} = 1 - L
\]

has the root 1 which is on and not outside the unit circle. That is the MA component is strictly noninvertible, and therefore the cointegrated model seems also that it cannot be approximated by autoregressive model.
In the next sections we study the handling of nonstationary bivariate autoregressions by estimating them in levels. There we discuss three ways (one way per section) of incorporating stochastic trend components into bivariate autoregressions.

**7.1 ADDITIVE MODEL**


This method focuses on generalizing the irregular term to a stationary autoregressive process. So then the bivariate model (in question) is taken as being the sum of a bivariate stochastic trend and a stationary bivariate autoregression. That is,

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \Phi(L)
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

(7.1.1a)

We take \( \Phi(L) = I - \Phi L \)

\[
\begin{bmatrix}
1 - \phi_{11} L & -\phi_{12} L \\
-\phi_{21} L & 1 - \phi_{22} L
\end{bmatrix}
\]

and (7.1.1a) becomes

\[
\begin{bmatrix}
1 - \phi_{11} L & -\phi_{12} L \\
-\phi_{21} L & 1 - \phi_{22} L
\end{bmatrix}
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

or

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix}
+ \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

(7.1.1a*)

If in (7.1.1) \( \Sigma_{\varepsilon \varepsilon} = 0 \), then \( \varepsilon_t = 0 \) both (7.1.1a) and (7.1.1b) give
\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mu_{1,t-1} + \beta_1 t + \eta_{1t} \\
\mu_{2,t-1} + \beta_2 t + \eta_{2t}
\end{bmatrix}
\]

which is a bivariate random walk with drift vector \( \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \).

By using (7.1.1b) to expand the above random walk with drift we obtain a detrended form

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\beta_{10} + \beta_1 t + w_{1t} \\
\beta_{20} + \beta_2 t + w_{2t}
\end{bmatrix}
\]

where

\[
\begin{bmatrix}
\beta_{10} \\
\beta_{20}
\end{bmatrix}
= \begin{bmatrix}
x_{1,0} \\
x_{2,0}
\end{bmatrix}
\]

and

\[
\begin{bmatrix}
w_{1t} \\
w_{2t}
\end{bmatrix}
= \sum_{j=0}^{t-1} \begin{bmatrix}
\eta_{1,t-j} \\
\eta_{2,t-j}
\end{bmatrix}
\]

If it is \( \Sigma_{\eta\eta} = 0 \), then

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mu_{1,0} + \beta_1 t \\
\mu_{2,0} + \beta_2 t
\end{bmatrix}
\]

which is still detrending, and (7.1.1a) becomes
and still a detrending. In both cases we get a continuous growth as $t$ increases, but none is a result of autoregression.

To investigate the autoregression, suppose there is no drift, that is

$$
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} = \begin{bmatrix} 0 \\
0 \end{bmatrix}
$$

then (7.1.1a*) leads us to the form

$$
\begin{bmatrix}
x_{1t} - \mu_{1t} \\
x_{2t} - \mu_{2t}
\end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix}
x_{1,t-1} - \mu_{1,t-1} \\
x_{2,t-1} - \mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\
\epsilon_{2t} \end{bmatrix}
$$

For the SSR we define

$$
\begin{bmatrix}
z_{1t}^{(1)} \\
z_{2t}^{(1)} \\
\cdots \\
z_{1t}^{(2)} \\
z_{2t}^{(2)}
\end{bmatrix} = \begin{bmatrix}
x_{1t} - \mu_{1t} \\
x_{2t} - \mu_{2t} \\
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
$$

and

$\epsilon_{t}^{(1)}$ and $\epsilon_{t}^{(2)}$ as bivariate white noise processes with common covariance matrix $I_2$.

The SSR becomes

$$
\begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix}
= \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix} \eta_{1t} \\
\eta_{2t} \end{bmatrix}
$$
We can show by induction, that for $k \in \mathbb{N}$, $k \geq 1,$

$$W^k = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^k \rightarrow O_2 \text{ as } h \rightarrow \infty,$$

that is, for large $k$

$$W^k \approx \begin{bmatrix} O_2 & : & O_2 \\ \vdots & \ddots & \vdots \\ O_2 & : & 1_2 \end{bmatrix}$$
\[
HW^k = \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^k \begin{bmatrix}
1 \\
0
\end{bmatrix}, \quad k = 1, 2, \ldots
\]

Let

\[
\begin{bmatrix}
z_1^{(1)}(t|t) \\
z_2^{(1)}(t|t) \\
\vdots \\
z_1^{(2)}(t|t) \\
z_2^{(2)}(t|t)
\end{bmatrix} =
\begin{bmatrix}
z_1^{(1)}(t|t) \\
z_2^{(1)}(t|t) \\
\vdots \\
z_1^{(2)}(t|t) \\
z_2^{(2)}(t|t)
\end{bmatrix}
\]

be given, then the \(\ell\)-step ahead forecasts are

\[
\begin{bmatrix}
x_1(t+\ell) \\
x_2(t+\ell)
\end{bmatrix} = HW^\ell z(t|t)
\]

\[
= \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}^\ell \begin{bmatrix}
z_1^{(1)}(t|t) \\
z_2^{(2)}(t|t)
\end{bmatrix} + \begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix}
\]

\[
\approx \begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix} \text{ for large } \ell
\]  

Given

\[
\Sigma_{zz}(t|t) = \begin{bmatrix}
\Sigma_{11}(t|t) & \Sigma_{12}(t|t) \\
\Sigma_{21}(t|t) & \Sigma_{22}(t|t)
\end{bmatrix}
\]

where

\[
\Sigma_{11}(t|t) = \text{cov}[z^{(1)}(t) - z^{(1)}(t|t)]
\]

\[
\Sigma_{22}(t|t) = \text{cov}[\mu(t) - \mu(t|t)]
\]

\[
\Sigma_{12}(t|t) = \text{cov}[z^{(1)}(t|t), \mu(t) - \mu(t|t)]
\]

\[
\Sigma_{21}(t|t) = \Sigma_{12}^t(t|t)
\]
We have
\[
BB' = \begin{bmatrix}
\Sigma_{\epsilon \epsilon} & 0 \\
0 & \Sigma_{\eta \eta}
\end{bmatrix}
\]
\[
\Sigma_{zz}(1) = W\Sigma_{zz}(t|t)W' + BB'
\]
\[
= \begin{bmatrix}
\Phi & 0 \\
0 & I_2
\end{bmatrix}
\begin{bmatrix}
\Sigma_{11}(t|t) & \Sigma_{12}(t|t)

\Sigma_{21}(t|t) & \Sigma_{22}(t|t)
\end{bmatrix}
\begin{bmatrix}
\Phi' \\
O
\end{bmatrix}
+ \begin{bmatrix}
\Sigma_{\epsilon \epsilon} & 0 \\
0 & \Sigma_{\eta \eta}
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\Phi \Sigma_{11}(t|t) \Phi' & \Phi \Sigma_{12}(t|t) \\
\Sigma_{21}(t|t) \Phi' & \Sigma_{22}(t|t)
\end{bmatrix}
+ \begin{bmatrix}
\Sigma_{\epsilon \epsilon} & 0 \\
0 & \Sigma_{\eta \eta}
\end{bmatrix}
\]
and for \( \ell \in \mathbb{N}, \ \ell \geq 2 \)
\[
\Sigma_{zz}(\ell) = W^\ell \Sigma_{zz}(t|t)(W^\ell)' + BB' + \sum_{j=1}^{\ell-1} W^j BB'(W^j)' + GG'
\]
\[
= \begin{bmatrix}
\Phi^\ell & 0 \\
0 & I_2
\end{bmatrix}
\begin{bmatrix}
\Sigma_{zz}(t|t) & (\Phi^\ell)' \\
0 & O
\end{bmatrix}
+ \begin{bmatrix}
\Sigma_{\epsilon \epsilon} & 0 \\
O & \Sigma_{\eta \eta}
\end{bmatrix}
\]
\[
+ \sum_{j=1}^{\ell-1} \begin{bmatrix}
\Phi^j & 0 \\
0 & I_2
\end{bmatrix}
\begin{bmatrix}
\Sigma_{\epsilon \epsilon} & 0 \\
0 & \Sigma_{\eta \eta}
\end{bmatrix}
\begin{bmatrix}
(\Phi^\ell)' \\
O
\end{bmatrix}
\]
Now
\[
\Sigma_{zz}(\ell) = H\Sigma_{zz}(\ell)H'
\]
\[
\ell = 1
\]
\[
\Sigma_{zz}(1) = \Phi \Sigma_{11}(t|t) \Phi' + \sum_{j=1}^{\ell} (\Phi^\ell)' + \sum_{j=1}^{\ell} (\Phi^\ell)' + \sum_{j=1}^{\ell} (\Phi^\ell)' + \sum_{j=1}^{\ell} (\Phi^\ell)'
\]
\[
\ell \in \mathbb{N}, \ \ell \geq 2
\]
\[
\Sigma_{zz}(\ell) = \Phi^\ell \Sigma_{11}(t|t)(\Phi^\ell)' + \sum_{j=1}^{\ell-1} (\Phi^\ell)'
\]
\[
+ \sum_{j=1}^{\ell-1} (\Phi^\ell)' + \sum_{j=1}^{\ell-1} (\Phi^\ell)'
\]
\[
+ \sum_{j=1}^{\ell-1} (\Phi^\ell)' + \sum_{j=1}^{\ell-1} (\Phi^\ell)'
\]
\[ = \Phi^\ell \Sigma_{11}(t|t)(\Phi^\ell)' + \Sigma_{21}(t|t)(\Phi^\ell)' + \Phi^\ell \Sigma_{12}(t|t) + \Sigma_{22}(t|t) + \ell \Sigma_{\eta \eta} + \Sigma_{\varepsilon \varepsilon} + \sum_{j=1}^{\ell-1} \Phi^j \Sigma_{\varepsilon \varepsilon} (\Phi^j)' \]  \tag{7.1.4a}

\[ \approx \Sigma_{22}(t|t) + \ell \Sigma_{\eta \eta} + \Sigma_{\varepsilon \varepsilon} + \sum_{j=1}^{\ell-1} \Phi^j \Sigma_{\varepsilon \varepsilon} (\Phi^j)' \]

for large \( \ell \): \( \Phi^\ell \to 0 \)  \tag{7.1.4b}

This forecast is linear on \( \ell \), and for a deterministic level, that is \( \Sigma_{\eta \eta} = 0 \), we have bounded MMSE, because by taking \( \Sigma_{\varepsilon \varepsilon} = I \)

\[ \sum_{j=1}^{\ell-1} \Phi^j \Sigma_{\varepsilon \varepsilon} (\Phi^j)' = \sum_{j=1}^{\ell-1} \Phi^j (\Phi^j)' = \sum_{j=1}^{\ell-1} (\Phi \Phi')^j \]

Now, using SDR, we write \( \Phi \Phi' = PD_{\lambda} P^{-1} \), where \( \lambda_1, \lambda_2 \) are squared eigenvalues of \( \Phi \), which are less than 1. Then

\[ \sum_{j=1}^{\ell-1} \lambda_1^j = \lambda_1 (1 - \lambda_1^{\ell-1}) / (1 - \lambda_1) = \lambda_1^* \text{ is convergent.} \]

Then

\[ \sum_{j=1}^{\ell-1} (\Phi \Phi')^j = PD_{\lambda^*} P^{-1} \]

The matrix \( \Phi \) dampens the contribution of error term \( \varepsilon \), and as seen in the point forecast, the effect of \( \Phi \) is not felt in the long run. As \( x_t - \mu(t|t) \) is a correction (error) added to correct \( \mu(t|t) \) in the point forecast \( x_t(\ell) \), this correction diminishes for large horizon \( \ell \), and is highest at \( \ell = 1 \).

### 7.2 BIVARIATE AUTOREGRESSIVE MODEL

Harvey (1989: 470-1)

The following model is an example of incorporating the bivariate autoregression within the (error of the) trend component. The model, modified for
bivariate case from Harvey (1989: 470 (8.8.7)), and after discussion with Prof. Markham, we have

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]  
(7.2.1a)

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \Phi^{-1}(L) \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]  
(7.2.1b)

We take

\[\Phi(L) = I - \Phi L\]

then (7.2.1b) becomes

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} - \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} - \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
\mu_{1,t-2} \\
\mu_{2,t-2}
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

which is

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
1 + \phi_{11} & \phi_{12} \\
\phi_{21} & 1 + \phi_{22}
\end{bmatrix} \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} - \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} \begin{bmatrix}
\mu_{1,t-2} \\
\mu_{2,t-2}
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]  
(7.2.1b*)

(This is a bivariate IAR(1,1) process.)

To derive SSR, with \( \varepsilon_t^{(1)} \) and \( \varepsilon_t^{(2)} \) as in previous sections, we define

\[
\begin{bmatrix}
z_{1t}^{(1)} \\
z_{2t}^{(1)} \\
\vdots \\
z_{1t}^{(2)} \\
z_{2t}^{(2)}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t} \\
\mu_{2t} \\
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix}
\]
\[
\begin{bmatrix}
1 + \phi_{11} & \phi_{12} & -\phi_{11} & -\phi_{12} \\
\phi_{21} & 1 + \phi_{22} & -\phi_{21} & -\phi_{22} \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\end{bmatrix}
+ 
\begin{bmatrix}
O_2 \\
\Sigma_{\eta}^{1/2} \eta \\
O_2 \\
O_2 \\
\end{bmatrix}
\begin{bmatrix}
e_1^{(1)} \\
e_2^{(1)} \\
e_1^{(2)} \\
e_2^{(2)} \\
\end{bmatrix}
\]

where again

\[
\begin{bmatrix}
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\end{bmatrix}
= \begin{bmatrix} 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{1,1}^{(1)} \\ z_{1,1}^{(2)} 
\end{bmatrix} \\
\begin{bmatrix} z_{2,1}^{(1)} \\ z_{2,1}^{(2)} 
\end{bmatrix} \\
\end{bmatrix}
+ \begin{bmatrix}
O_2 \\
\Sigma_{\Sigma_\epsilon}^{1/2} \Sigma_\epsilon \\
O_2 \\
O_2 \\
\end{bmatrix}
\begin{bmatrix}
e_1^{(1)} \\
e_2^{(1)} \\
e_1^{(2)} \\
e_2^{(2)} \\
\end{bmatrix}
\]

We define

\[
W_0 = I_4 \quad \text{and} \quad \phi^k = O_2 \quad \text{for} \quad k < 0
\]

and we can show by induction that for \( k \in \mathbb{N}, \ k \geq 2 \)

\[
W^k = \begin{bmatrix}
\sum_{j=0}^{k} \phi^j & -\phi \sum_{j=0}^{k-1} \phi^j \\
\sum_{j=0}^{k-1} \phi^j & -\phi \sum_{j=0}^{k-2} \phi^j \\
\end{bmatrix}
= (I - \phi)^{-1} \begin{bmatrix}
I - \phi^{k+1} & -(I - \phi^k) \phi \\
I - \phi^k & -(I - \phi^{k-1}) \phi \\
\end{bmatrix}
\]

\[
\approx (I - \phi)^{-1} \begin{bmatrix} I & -\phi \\
I & -\phi \\
\end{bmatrix} \quad \text{for large} \quad k
\]
Given

\[
\begin{bmatrix}
z_1^{(1)}(t|t) \\ z_2^{(1)}(t|t) \\ \vdots \\ z_1^{(2)}(t|t) \\ z_2^{(2)}(t|t)
\end{bmatrix} =
\begin{bmatrix}
\mu_1(t|t) \\ \mu_2(t|t) \\ \mu_1(t-1|t) \\ \mu_2(t-1|t)
\end{bmatrix}
\]

then

\[
\begin{bmatrix}
x_{1t}(\ell) \\ x_{2t}(\ell)
\end{bmatrix} = HWz(t|t)
\]

\[
= (I - \Phi)^{-1}[(I - \Phi^{\ell+1})z^{(1)}(t|t) - (I - \Phi^{\ell})\Phi z^{(2)}(t|t)]
\]  
(7.2.2a)

\[
\approx (I - \Phi)^{-1}[\mu(t|t) - \Phi \mu(t-1|t)] 	ext{ for large } \ell
\]  
(7.2.2b)

and

\[
\begin{bmatrix}
x_{1t}^{(1)} \\ x_{2t}^{(1)}
\end{bmatrix} = HWz(t|t)
\]

\[
= (I + \Phi)\mu(t|t) - \Phi \mu(t-1|t)
\]

Suppose we are given

\[
\Sigma_{zz}(t|t) =
\begin{bmatrix}
\Sigma_{11}(t|t) & \Sigma_{12}(t|t) \\ \Sigma_{21}(t|t) & \Sigma_{22}(t|t)
\end{bmatrix}
\]

then with

\[
BB' =
\begin{bmatrix}
\Sigma \eta \eta & O_2 \\ O_2 & O_2
\end{bmatrix}
\]

\[
\Sigma_{zz}(1) = [I + \Phi - \Phi]\Sigma_{zz}(t|t)[I + \Phi' \ I] + \Sigma \eta \eta \ O'
\]

and for \( \ell \in N, \ell \geq 2 \)
168

\[ \Sigma_{zz}(t) = W^\ell \Sigma_{zz}(t|t)(W^\ell)' + \sum_{j=0}^{\ell-1} W^j \begin{bmatrix} \Sigma \eta \eta & O \\ O & O \end{bmatrix} (W^j)' \]

Then

\[ \Sigma_{zz}(1) = \begin{bmatrix} I & O \end{bmatrix} \Sigma_{zz}(1) \begin{bmatrix} I & O \end{bmatrix}' + GG' \]

\[ = (I+\Phi)\Sigma_{11}(t|t)(I+\Phi)' - (I+\Phi)\Sigma_{12}(t|t)\Phi' \]

\[ - \Phi \Sigma_{21}(t|t)(I+\Phi)' + \Phi \Sigma_{22}(t|t)\Phi' + \Sigma \eta \eta + \Sigma \epsilon \epsilon \quad (7.2.3a) \]

and for \( \ell \in \mathbb{N}, \ell \geq 2 \)

\[ \Sigma_{zz}(\ell) = HW^\ell \Sigma_{zz}(t|t)(HW^\ell)' + \sum_{j=0}^{\ell-1} HW^j \begin{bmatrix} \Sigma \eta \eta & O \\ O & O \end{bmatrix} (HW^j)' + GG' \]

\[ = (I-\Phi)^{-1} \left\{ \Sigma_{11}(t|t) - \Phi \Sigma_{21}(t|t) - \Sigma_{12}(t|t)\Phi' \right\} (I-\Phi')^{-1} \]

\[ + \Phi \Sigma_{22}(t|t)\Phi' \right\} (I-\Phi')^{-1} \]

\[ + (I-\Phi)^{-1} \left\{ \Sigma \eta \eta - \Phi j+1 \Sigma \eta \eta - \Sigma \eta \eta (\Phi')^{j+1} \right\} (I-\Phi')^{-1} + \Sigma \epsilon \epsilon + \Sigma \eta \eta \]

where we assumed \( \Phi \ell \approx 0 \)

\[ = (I-\Phi)^{-1} \left\{ \Sigma_{11}(t|t) - \Phi \Sigma_{21}(t|t) - \Sigma_{12}(t|t)\Phi' \right\} 

+ \Phi \Sigma_{22}(t|t)\Phi' \right\} (I-\Phi')^{-1} + (\ell-1)(I-\Phi)^{-1} \Sigma \eta \eta (I-\Phi')^{-1} \]

\[ - (I-\Phi)^{-2} \Phi \Sigma \eta \eta (I-\Phi')^{-1} - (I-\Phi)^{-1} \Sigma \eta \eta (\Phi')^2 (I-\Phi')^{-2} \]

\[ + (I-\Phi)^{-1} \left\{ \sum_{j=1}^{\ell-1} \Phi j+1 \Sigma \eta \eta (\Phi')^{j+1} (I-\Phi')^{-1} + \Sigma \epsilon \epsilon + \Sigma \eta \eta \quad (7.2.3b) \right\} \]

and this is linear on \( \ell \).
The MMSE is damped at many parts by $\Phi$ and we do not see terms such as $t^2$. Thus, the expansion is not quick, and this was displayed by the additive model as well.

7.3 DYNAMIC MODEL


We set up an autoregressive model by including $I - \Phi L$ on the left hand side as usual, then we have

$$
\begin{bmatrix}
1 - \phi_{11} L & -\phi_{12} L \\
-\phi_{21} L & 1 - \phi_{22} L
\end{bmatrix}
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix}
= \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
$$

(7.3.1a)

$$
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
$$

(7.3.1b)

The trend is a bivariate random walk plus noise with drift so that each $\mu_{it} \sim I(1)$, and also each $x_{it}$ is nonstationary but

$$
(I - \Phi L) \begin{bmatrix}
\Delta x_{1t} \\
\Delta x_{2t}
\end{bmatrix} = \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} + \varepsilon_{1t} - \varepsilon_{1,t-1} \\
\eta_{2t} + \varepsilon_{2t} - \varepsilon_{2,t-1}
\end{bmatrix}
$$

is stationary, then each $x_{it} \sim I(1)$, $i = 1, 2$.

When $\Sigma_{\eta\eta} = 0$, we have a deterministic trend so that

$$
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,0} \\
\mu_{2,0}
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} t
$$

which is detrending, and

$$
(I - \Phi L) \begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,0} \\
\mu_{2,0}
\end{bmatrix} + \begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} t + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
$$
If $\Sigma_{ee} = 0$ then

$$(I - \Phi L)[x_{1t}] = [\mu_{1t}]
[x_{2t}] = [\mu_{2t}]$$

$$= [\mu_{1,t-1}] + [\beta_{1}] + [\eta_{1t}]
[\mu_{2,t-1}] [\beta_{2}] [\eta_{2t}]$$

$$= [\mu_{1,0}] + [\beta_{1}] t + [\eta^{*}_{1t}]
[\mu_{2,0}] [\beta_{2}] [\eta^{*}_{2t}]$$

where

$$[\eta^{*}_{1t}] = \sum_{i=0}^{t-1} [\eta_{1,t-j}]
[\eta^{*}_{2t}] = [\eta_{2,t-j}]$$

This means that if any of the errors is removed, we obtain a detrending for the slope.

Writing (7.3.1) in another form and assuming $\beta = 0$, then

$$[x_{1t}] = [\mu_{1t}] + [\phi_{11} \phi_{12}] [x_{1,t-1} - \mu_{1,t-1}] + [\phi_{11} \phi_{12}] [\mu_{1,t-1}]
[x_{2t}] = [\mu_{2t}] + [\phi_{21} \phi_{22}] [x_{2,t-1} - \mu_{2,t-1}] + [\phi_{21} \phi_{22}] [\mu_{2,t-1}]$$

$$+ [\varepsilon_{1t} \varepsilon_{2t}]$$

(7.3.1a*)

$$[\mu_{1t}] = [\mu_{1,t-1}] + [\eta_{1t}]
[\mu_{2t}] = [\mu_{2,t-1}] + [\eta_{2t}]$$

(7.3.1b*)

where in the first equation we have substituted $[\mu_{1,t-1}]$ for convenience of deriving SSR which we are about to do. Comparing (7.3.1a*) with (7.1.1a*)
of additive model (which is of course bivariate stationarity AR(1)), the
dynamic model is more by the term \( \Phi \mu_{t-1} \).

The SSR is

\[
\begin{bmatrix}
z_{1t}^{(1)} \\
z_{2t}^{(1)} \\
\vdots \\
z_{1t}^{(2)} \\
z_{2t}^{(2)}
\end{bmatrix} =
\begin{bmatrix}
x_{1t} - \mu_{1t} \\
x_{2t} - \mu_{2t} \\
\vdots \\
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
z_{1t}^{(1)} \\
z_{2t}^{(1)} \\
z_{1t}^{(2)} \\
z_{2t}^{(2)}
\end{bmatrix} +
\begin{bmatrix}
\frac{1}{\epsilon_e} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
\varepsilon_t^{(1)} \\
\varepsilon_t^{(2)}
\end{bmatrix}
\]

\[
\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
z_{1t}^{(1)} \\
z_{2t}^{(1)} \\
\vdots \\
z_{1t}^{(2)} \\
z_{2t}^{(2)}
\end{bmatrix}
\]

We define

\[ W^0 = I_4 \]

and we can show by induction that for \( k \in \mathbb{N}, \ k \geq 1 \)

\[
W^k = \begin{bmatrix}
\phi_k & \sum_{j=1}^{k} \phi_j \\
0 & I
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\phi_k & \phi(I-\Phi)^{-1}(I-\Phi^k) \\
0 & I
\end{bmatrix}
\]
\[
\begin{bmatrix}
O & \Phi(I-\Phi)^{-1} \\
O & I
\end{bmatrix}
\] for large \( k \)

\[
HW_k = [\Phi^k \sum_{j=0}^k \Phi^j], \quad k = 1, 2, \ldots
\]

\[
\approx [O \quad (I-\Phi)^{-1}] \quad \text{for large } k
\]

Let
\[
\begin{bmatrix}
z_1^{(1)}(t|t) \\
z_2^{(1)}(t|t) \\
\vdots \\
z_1^{(2)}(t|t) \\
z_2^{(2)}(t|t)
\end{bmatrix} =
\begin{bmatrix}
x_1(t|t) - \mu_1(t|t) \\
x_2(t|t) - \mu_2(t|t) \\
\vdots \\
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix}
\]

then the \( \ell \)-step ahead forecasts are given by
\[
\begin{bmatrix}
x_{1t}(\ell) \\
x_{2t}(\ell)
\end{bmatrix} = HW_\ell z(t|t)
\]

\[
= \Phi^\ell z^{(1)}(t|t) + \sum_{j=0}^\ell \Phi^j z^{(2)}(t|t)
\]

(7.3.2a)

\[
\approx (I-\Phi)^{-1}\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix}
\] for large \( \ell \)

(7.3.2b)

Also, given
\[
\Sigma_{zz}(t|t) = \begin{bmatrix}
\Sigma_{11}(t|t) & \Sigma_{12}(t|t) \\
\Sigma_{21}(t|t) & \Sigma_{22}(t|t)
\end{bmatrix}
\]

then, evaluated only for large \( \ell \),
\[ \Sigma_{xx}(\ell) = (I-\Phi)^{-1}\Sigma_{22}(t|t)(I-\Phi')^{-1} + \sum_{e} + \sum_{\eta} + \sum_{j=1}^{\ell-1} \Phi^j \epsilon^e (\Phi')^j \]

\[ + (\ell-1)(I-\Phi)^{-1}\eta \eta^e (I-\Phi')^{-1} - (I-\Phi)^{-2}\eta^2 \eta (I-\Phi')^{-1} \]

\[ - (I-\Phi)^{-1}\sum_{\eta} (\Phi')^2 (I-\Phi')^{-2} \]

\[ + (I-\Phi)^{-1}\sum_{j=1}^{\ell-1} \Phi^j \eta \eta^e (\Phi')^j (I-\Phi')^{-1} \]

(7.3.3)

This is also linear on \( \ell \) on one term, and is dominated by damping factors \( \Phi \). The expansion is relatively slow. The point forecasts do not get unbounded, in fact they get more damped in the long run.

7.4 CLOSING REMARKS

The additive model and the vector autoregression have hidden features, seemingly because the error terms hides the structure inside the autoregression. Engle & Yoo (1987: 143) argue that autoregressions require certain restrictions before they can perform well in forecasting, and on page 158 these authors conclude that "vector autoregressive model in differences is inappropriate because model suffers misspecification and forecasts will diverge from each other".

In the dynamic model, the autoregression has always been visible. Unlike in the other two models, deterministic behavior does not wipe off the autoregression.
CHAPTER 8

COMMON LEVELS/TRENDS AND
COINTEGRATION OF AUTOREGRESSIONS

In the previous chapter, bivariate autoregressions were discussed, and in Chapter 6 we discussed dynamic factor analysis where we had some of the bivariate components having a common factor. These common factors gave a good account in the description of the concept of cointegration. Any bivariate random component (parameter) which has a common factor will be replaced by one random common parameter which will be loaded by nonrandom components, both nonzero.

This chapter is a continuation of discussions of autoregressive processes of the previous chapter, the dynamic factor analysis and cointegration concepts as our focus. The sequence followed in Chapter 7 for the three approaches discussed, will be followed in the forthcoming discussion:

8.1 ADDITIVE MODEL


We introduce a common trend for the autoregressive additive model of (7.1.1), and the loading vector \( \theta = [\theta_1 \theta_2]^\top \) will be taken for convenience as \( \theta = [1 \ 0]^\top \), where \( \theta \neq 0 \) (as we said in the first paragraph of this chapter that both \( \theta_1 \neq 0 \) and \( \theta_2 \neq 0 \)). We now have the model, also given in Harvey (1989: (8.5.5)):

\[
\begin{bmatrix}
\tau_{1t} \\
\tau_{2t}
\end{bmatrix} = \begin{bmatrix}
1 \\
\theta
\end{bmatrix} \mu_t + \Phi^{-1}(L) \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

\[
\mu_t = \mu_{t-1} + \beta + w_t
\]
where $w_t$ and $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ are uncorrelated white noise processes.

We already know that $\mu_t$ (random walk plus noise with drift) is nonstationary, and that

$$\Delta \mu_t = w_t^*$$

is stationary, where $w_t^* = w_t - w_{t-1}$.

Because each $x_{1t}$ is linear on $\mu_t$ and error terms, these components (the $x_{1t}$) are nonstationary as well, and

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ \theta \end{bmatrix} w_t^* + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

each component of which is stationary, where $\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = \Phi^{-1}(L) \begin{bmatrix} \Delta \varepsilon_{1t} \\ \Delta \varepsilon_{2t} \end{bmatrix}$.

That is $x_{1t}, x_{2t} \sim I(1)$.

Let $\alpha = \begin{bmatrix} 1 & -1/\theta \end{bmatrix}'$, then

$$\alpha^\prime \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -1/\theta \end{bmatrix} \Phi^{-1}(L) \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

(8.1.2)

The right-hand side is stationary because the stationary white noise is multiplied by polynomials in $L$, and therefore (because $\alpha^\prime x_t$ is also stationary by equality), $\alpha$ is a cointegrating vector.

Therefore, (8.1.1) is a cointegrated process with a cointegrating vector.
By writing \( \Phi^{-1}(L) \) as:

\[
\Phi^{-1}(L) = \sum_{j=0}^{\infty} \begin{bmatrix} \phi_{11,j}L^j & \phi_{12,j}L^j \\ \phi_{21,j}L^j & \phi_{22,j}L^j \end{bmatrix}
\]

and setting the white noise process \( \eta_t \) as

\[
\eta_t = \sum_{j=0}^{\infty} \left[ \phi_{11,j}L^j - \left(1/\theta \right)\phi_{12,j} \right] \varepsilon_{1t} + \sum_{j=0}^{\infty} \left[ \phi_{12,j} - \left(1/\theta \right)\phi_{22,j} \right] \varepsilon_{2t},
\]

the cointegrating relationship \((8.1.2)\) can be rewritten as

\[
x_{1t} - \left(1/\theta \right)x_{2t} = \eta_t
\]

which shows that there is a levels relationship between \( x_{1t} \) and \( x_{2t} \).

In \((8.1.1)\) the nonstationary component is \( \mu_t \) with loading vector \( \theta = \begin{bmatrix} 1 \\ \theta \end{bmatrix} \).

The cointegrating vector \( \alpha \) is such that

\[
\alpha^t \theta = 0. \quad (8.1.3)
\]

**8.2 BIVARIATE AUTOREGRESSIVE TRENDS**


We assume that the bivariate autoregressive trend model \((7.2.1)\) has a common level, this level is a random walk and it is the noise of this random walk that has an autoregression. That is,
where the white noise processes $[\varepsilon_{1t}, \varepsilon_{2t}]$ and $w_t$ are uncorrelated.

The random walk (univariate) process (8.2.1c) is nonstationary, and

$$\Delta \mu_t = \phi^{-1}(L)w_t = w_t^*$$

is stationary, and this leads to

$$\begin{bmatrix} \Delta \mu_{1t} \\ \Delta \mu_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_t^* + \begin{bmatrix} \eta_{1t} - \eta_{1,t-1} \\ \eta_{2t} - \eta_{2,t-1} \end{bmatrix}$$

which is also stationary, and lastly

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_t^* + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}$$

which is stationary, where

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} \eta_{1t} - \eta_{1,t-1} \\ \eta_{2t} - \eta_{2,t-1} \end{bmatrix} + \begin{bmatrix} \Delta \varepsilon_{1t} \\ \Delta \varepsilon_{2t} \end{bmatrix}$$

is a bivariate white noise. That is, $x_{1t} \sim I(1)$ and $x_{2t} \sim I(1)$.

By writing (by substituting (8.2.1b) into (8.2.1a),

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

$$\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mu_t + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}$$

$$\mu_t = \mu_{t-1} + \phi^{-1}(L)w_t$$
where the white noise process \( v_t^* \) is

\[
\begin{bmatrix}
  v_{1t}^* \\
  v_{2t}^*
\end{bmatrix} =
\begin{bmatrix}
  \eta_{1t} \\
  \eta_{2t}
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_{1t} \\
  \varepsilon_{2t}
\end{bmatrix}
\]

then, with \( \alpha = \begin{bmatrix} 1 & -1/	heta \end{bmatrix} \), then

\[
\alpha' \begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} = \alpha' \begin{bmatrix}
  v_{1t}^* \\
  v_{2t}^*
\end{bmatrix}
\]

is stationary. Therefore

\[
\begin{bmatrix}
  x_{1t} \\
  x_{2t}
\end{bmatrix} \sim CI(1,1)
\]

with cointegration vector \( \alpha \).

By defining \( v_t = v_{1t}^* - (1/\theta)v_{2t}^* \), then (8.2.2) may be written as

\[
x_{1t} = (1/\theta)x_{2t} + v_t
\]

and there is (again) a levels relationship between \( x_{1t} \) and \( x_{2t} \). Also, with the loading vector \( \theta = \begin{bmatrix} 1 & \theta \end{bmatrix} \), the relationship between \( \theta \) and the cointegrating vector \( \alpha \) is such that

\[
\alpha' \theta = 0.
\]

8.3 DYNAMIC MODEL

The autoregressive (dynamic) model (7.3.1) is recalled and we introduce a
common trend $\mu_t$. Then a new form is given by

$$\Phi(L)[x_{1t}] = [1] \mu_t + [0] + [\varepsilon_{1t}]$$  \hspace{1cm} (8.3.1a)

$$\mu_t = \mu_{t-1} + \beta + \eta_t$$  \hspace{1cm} (8.3.1b)

as the trend has reduced to univariate because $\mu_{1t} = \mu_{2t}$.

The random walk with drift (8.3.1b) is nonstationary as before, with

$$\Delta \mu_t = \eta_t - \eta_{t-1} = \eta_t^*$$

a stationary process. Therefore (8.3.1a) as a linear function of $\mu_t$, it is nonstationary and

$$\Phi(L)[\Delta x_{1t}] = [1] \eta_t^* + [\varepsilon_{1t} - \varepsilon_{1,t-1}]$$

$$\Delta x_{2t} = [0] [\varepsilon_{2t} \varepsilon_{2,t-1}]$$

is a stationary process, so that we can conclude that $x_{1t} \sim I(1)$ and $x_{2t} \sim I(1)$.

Let $\Phi(L)$ be of order $p$, so we can write as

$$\Phi(L) = \left[ \sum_{j=0}^{p} \phi_{11,j} L^j \sum_{j=0}^{p} \phi_{12,j} L^j \right]$$

$$\Phi(0) = I_2$$

and we assume that all the roots of $\Phi(z)$ lie outside the unit circle. Then $\Phi(1)$ is nonsingular (roots lie outside unit circle), and we can write $\Phi(L)$ as
\[ \Phi(L) = \Phi(1)L + \Phi^*(L)(1-L) \]  
\begin{align*}
\Phi^*(L) &= \left[ \sum_{j=0}^{p-1} \phi_{11,j}^* L^j, \sum_{j=0}^{p-1} \phi_{12,j}^* L^j \right] \\
&\quad \left[ \sum_{j=0}^{p-1} \phi_{21,j}^* L^j, \sum_{j=0}^{p-1} \phi_{22,j}^* L^j \right]
\end{align*}

is of order \( p-1 \), and
\[ \Phi^*(0) = I_2 \]

Substituting the right-hand side of (8.3.2) for \( \Phi(L) \) in (8.3.1a), recalling that \( \Delta = 1 - L \), and rearranging we obtain

\[ \Phi^*(L) \Delta \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = -\Phi(1)L \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} + \begin{bmatrix} 1 \mu_t + 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \]

which in long-hand notation is

\[ \begin{bmatrix}
\sum_{j=0}^{p-1} \phi_{11,j}^* L^j \\
\sum_{j=0}^{p-1} \phi_{12,j}^* L^j
\end{bmatrix} \begin{bmatrix} \Delta z_{1t} \\ \Delta z_{2t} \end{bmatrix} = -\begin{bmatrix}
\sum_{j=0}^{p} \phi_{11,j} \\
\sum_{j=0}^{p} \phi_{12,j}
\end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \mu_t + 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \]

The inverse of \( \Phi(1) \) is
\[ \Phi^{-1}(1) = \frac{1}{D} \begin{bmatrix} \Sigma_{\phi_{22,j}}^{p} & -\Sigma_{\phi_{12,j}}^{p} \\
\Sigma_{\phi_{21,j}}^{p} & \Sigma_{\phi_{11,j}}^{p} \end{bmatrix} \]  

where

\[ D = \Sigma_{\phi_{11,j}}^{p} \Sigma_{\phi_{22,j}}^{p} - \Sigma_{\phi_{12,j}}^{p} \Sigma_{\phi_{21,j}}^{p} \]

We multiply the equation (8.3.3) by (8.3.4) to obtain

\[ \begin{bmatrix} \Sigma_{\phi_{22,j}}^{p} & -\Sigma_{\phi_{12,j}}^{p} \\
\Sigma_{\phi_{21,j}}^{p} & \Sigma_{\phi_{11,j}}^{p} \end{bmatrix} \begin{bmatrix} \Sigma_{\phi_{11,j}}^{p} L_j^{j=0} & \Sigma_{\phi_{12,j}}^{p} L_j^{j=0} \\
\Sigma_{\phi_{21,j}}^{p} L_j^{j=0} & \Sigma_{\phi_{22,j}}^{p} L_j^{j=0} \end{bmatrix} \begin{bmatrix} \Delta x_{1t} \\
\Delta x_{2t} \end{bmatrix} \]

\[ = -\begin{bmatrix} x_{1,t-1} \\
x_{2,t-1} \end{bmatrix} + \frac{1}{D} \begin{bmatrix} \Sigma_{\phi_{22,j}}^{p} & -\Sigma_{\phi_{12,j}}^{p} \\
\Sigma_{\phi_{21,j}}^{p} & \Sigma_{\phi_{11,j}}^{p} \end{bmatrix} \begin{bmatrix} \Delta x_{1t} \\
\Delta x_{2t} \end{bmatrix} \]

\[ + \frac{1}{D} \begin{bmatrix} \Sigma_{\phi_{12,j}}^{p} \\
\Sigma_{\phi_{11,j}}^{p} \end{bmatrix} \mu_t + \frac{1}{D} \begin{bmatrix} \Sigma_{\phi_{22,j}}^{p} & -\Sigma_{\phi_{12,j}}^{p} \\
\Sigma_{\phi_{21,j}}^{p} & \Sigma_{\phi_{11,j}}^{p} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\
\epsilon_{2t} \end{bmatrix} \]

(8.3.5)

In (8.3.1) the loading vector \( \Theta = \begin{bmatrix} 1 \end{bmatrix} \) has only one element \( \theta \) which is unknown and this implies (heuristically) that there need only be one unknown element in the cointegrating vector \( \alpha = \begin{bmatrix} \alpha \\
1 \end{bmatrix} \) (or \( \begin{bmatrix} 1 \end{bmatrix} \)).

The following result was proved with the help of and after discussion with Prof. Markham.
Result 8.1

Let \[ \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \] be the autoregressive dynamic model with common trend \( \mu_t \) and loading vector \( \theta = [1] \) as in (8.3.1)/(8.3.5). Then

(1) A necessary and sufficient condition for \( \alpha = \begin{bmatrix} \alpha \\ 1 \end{bmatrix} \) to be a cointegrating vector is that

\[
\alpha \sum_{j=0}^{p} \phi_{22,j} + \theta \sum_{j=0}^{p} \phi_{11,j} = \alpha \theta \sum_{j=0}^{p} \phi_{12,j} + \sum_{j=0}^{p} \phi_{21,j} \quad (8.3.6)
\]

and

\[
\alpha = \frac{\sum_{j=0}^{p} \phi_{21,j} - \theta \sum_{j=0}^{p} \phi_{11,j}}{\sum_{j=0}^{p} \phi_{22,j} - \theta \sum_{j=0}^{p} \phi_{12,j}} \quad (8.3.7)
\]

Before we provide proof of this result we discuss its existence under the given conditions. It is clear that \( \alpha \) is defined only if \( \sum_{j=0}^{p} \phi_{22,j} - \theta \sum_{j=0}^{p} \phi_{12,j} \neq 0 \).

This is true because if not, then

\[
\sum_{j=0}^{p} \phi_{22,j} = \theta \sum_{j=0}^{p} \phi_{12,j} \quad (8.3.8a)
\]

Substituting in (8.3.6) and simplifying we have

\[
\theta \sum_{j=0}^{p} \phi_{11,j} = \sum_{j=0}^{p} \phi_{21,j} \quad (8.3.8b)
\]

These two imply that
and this implies that $\Phi^{-1}(1)$ in (8.3.4) is singular, which contradicts the existence of $\Phi(1) = [\Phi^{-1}(1)]^{-1}$. Therefore, under the given conditions, (8.3.7) is possible, we only need to prove (8.3.6) and derive $\alpha$ of (8.3.7) by making $\alpha$ subject of the formula in (8.3.6).

Proof of Result 8.1

The common factor $\mu_t$ is the nonstationary component which makes the whole process (8.3.1) nonstationary. The cointegrating vector removes the nonstationary component, hence by multiplying by $\alpha$ at equation (8.3.5), to ensure that the coefficient of $\mu_t$ is 0, then $\alpha$ is a cointegrating vector if and only if

$$\alpha' \Phi^{-1}(1) \theta = 0$$  \hspace{1cm} (8.3.9)

But

$$\alpha' \Phi^{-1}(1) \theta = \frac{1}{D} [\alpha \begin{bmatrix} \sum_{j=0}^{p} \phi_{22,j} & - \sum_{j=0}^{p} \phi_{12,j} \\ - \sum_{j=0}^{p} \phi_{21,j} & \sum_{j=0}^{p} \phi_{11,j} \end{bmatrix} \begin{bmatrix} \theta \end{bmatrix}]$$

$$= \frac{1}{D} \left[ \alpha \sum_{j=0}^{p} \phi_{22,j} - \alpha \theta \sum_{j=0}^{p} \phi_{12,j} - \sum_{j=0}^{p} \phi_{21,j} + \theta \sum_{j=0}^{p} \phi_{11,j} \right]$$  \hspace{1cm} (8.3.10)

Substituting (8.3.9) into (8.3.10) and rearranging the result leads to

$$\alpha \sum_{j=0}^{p} \phi_{22,j} + \theta \sum_{j=0}^{p} \phi_{11,j} = \alpha \theta \sum_{j=0}^{p} \phi_{12,j} + \sum_{j=0}^{p} \phi_{21,j}$$

which is (8.3.6). This may be written as

$$\alpha \left[ \sum_{j=0}^{p} \phi_{22,j} - \theta \sum_{j=0}^{p} \phi_{12,j} \right] = \sum_{j=0}^{p} \phi_{21,j} - \theta \sum_{j=0}^{p} \phi_{11,j}$$
so that

\[
\alpha = \frac{\sum_{j=0}^{p} \phi_{21,j} - \theta \sum_{j=0}^{p} \phi_{11,j}}{\sum_{j=0}^{p} \phi_{22,j} - \theta \sum_{j=0}^{p} \phi_{12,j}}
\]

which is (8.3.7).

Now (8.3.6) is valid only if (8.3.9) is valid. If (8.3.9) is not true for all \( \alpha \), then premultiplying by any vector, (8.3.5) will contain \( \mu_t \) and therefore will always be nonstationary. That is, there is no cointegration.

Conversely, to have cointegration is to free the equation from \( \mu_t \) when multiplying by a cointegration vector \( \alpha' \), that is (8.3.6) must be true.

Let \( p = 1 \), then

\[
\Phi(L) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix} L,
\]

and condition (8.3.6) for \( \alpha \) to be a cointegrating vector reduces to

\[
\alpha(1-\phi_{22}) + \theta(1-\phi_{11}) = -\alpha \phi_{12} - \phi_{21}
\]

then

\[
\alpha = -\frac{\phi_{21} + \theta(1-\phi_{11})}{1 - \phi_{22} + \theta \phi_{12}}
\]

Then the cointegrating vectors may take the form

\[
\alpha^* = \begin{bmatrix} 1 - \phi_{22,j} + \theta \phi_{12,j} \\ - \phi_{21,j} + \theta(\phi_{11,j} - 1) \end{bmatrix}
\]

Suppose that (8.3.1) is SUTSE, then \( \phi_{12,j} = \phi_{21,j} = 0 \) for all \( j = 0, 1, \ldots, p \), then the cointegrating condition (8.3.6) reduces to
\[
\alpha \sum_{j=0}^{p} \phi_{22,j} + \theta \sum_{j=0}^{p} \phi_{11,j} = 0
\]

and from (8.3.7) then we obtain

\[
\alpha = -\sum_{j=0}^{p} \phi_{11,j}
\]

\[
\sum_{j=0}^{p} \phi_{22,j}
\]

When \( p = 1 \), we have as usual

\[
\Phi(L) = \begin{bmatrix} 1-\phi_{11}L & 0 \\ 0 & 1-\phi_{22}L \end{bmatrix}
\]

and the cointegrating relation is

\[
\alpha(1-\phi_{22}) + \theta(1-\phi_{11}) = 0
\]

and because

\[
\Phi(1) = \begin{bmatrix} 1-\phi_{11} & 0 \\ 0 & 1-\phi_{22} \end{bmatrix}
\]

is nonsingular \( \phi_{11} \neq 1 \) and \( \phi_{22} \neq 1 \) and then \( \alpha \) becomes an integrating vector if

\[
\alpha = -\frac{\theta(1-\phi_{11})}{1-\phi_{22}}
\]

At the beginning we proved that \( z_{1t} \sim I(1) \), \( z_{2t} \sim I(1) \). Now if \( \Phi(1) \) is nonsingular, then by Result 8.1
\[
\begin{bmatrix}
  z_{1t} \\
  z_{2t}
\end{bmatrix}
\sim CI(1,1)
\]

The results \( \alpha' \theta = 0 \) of (8.1.3) and (8.2.3) are special cases of Result 8.1 because \( \Phi(L) = I_2 \).

In all the conditions above, the cointegrating vector always depend on the elements of the autoregressive matrix and the loading vector of a common factor. The autoregressions for a dynamic model lead to easier interpretations because in the two previous processes discussed, we could not involve the autoregressions for the cointegrating vectors.

8.4 REMARK


Cointegration is understood well when there is a common nonstationary component for the two series. In the vector autoregressive models the common trends may be introduced to wipe off the autoregression when the cointegration relation \( \alpha' \Phi(1) \theta = 0 \) is applied. In all the three models considered, the presence of common trends means that the series are cointegrated.

The autoregressive polynomial \( \Phi(L) \) when written as,

\[
\Phi(L) = \Phi(1)L + \Phi^*(L)\Delta
\]

is usually called Granger's representation. This form is so helpful that many articles (such as Engle & Yoo (1987: 143-159) and Davidson (1991: 41-62)) consider it fundamental in cointegration. This is the form we used to derive Result 8.1.

Cointegration is simply in terms of the elements of the coefficient matrix (or vector) of the common factor. The inclusion of common factors in all the work were applications of dynamic factor analysis and this is irrespective of whether such application was consciously done or not. We conclude that dynamic factor analysis is a requirement for existence of cointegration relationship.
When we discussed forecasting for the ARIMA models, we said some of the (important and) desirable properties of forecasts are accuracy and flexibility. This is because forecasts are often used to develop further forecasts. Because of this role, forecasts should be reliable, and this reliability may be verified statistically using diagnostics or statistical tests. Also, when considered to develop other forecasts, the current set of forecasts must be valid (up-to-date and accurate). We use the SSR and the Kalman filter to develop forecasts for the bivariate structural time models which were introduced in Chapter 2 as generalizations of certain specific bivariate exponential smoothing algorithms.

9.1 POINT FORECasts AND FORECAST MSE'S

9.1.1 Random walk plus noise


The random walk plus noise is given by

\[
\begin{bmatrix}
    x_{1t} \\
    x_{2t}
\end{bmatrix} =
\begin{bmatrix}
    \mu_{1t} \\
    \mu_{2t}
\end{bmatrix} +
\begin{bmatrix}
    \epsilon_{1t} \\
    \epsilon_{2t}
\end{bmatrix} \tag{9.1.1a}
\]

\[
\begin{bmatrix}
    \mu_{1t} \\
    \mu_{2t}
\end{bmatrix} =
\begin{bmatrix}
    \mu_{1,t-1} \\
    \mu_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
    \eta_{1t} \\
    \eta_{2t}
\end{bmatrix} \tag{9.1.1b}
\]

where \( \epsilon_{1t} \) and \( \epsilon_{2t} \) are uncorrelated white noise processes with
covariance matrices \[
\begin{bmatrix}
\sigma_{11, \varepsilon} & \sigma_{12, \varepsilon} \\
\sigma_{21, \varepsilon} & \sigma_{22, \varepsilon}
\end{bmatrix}
\] and \[
\begin{bmatrix}
\sigma_{11, \eta} & \sigma_{12, \eta} \\
\sigma_{21, \eta} & \sigma_{22, \eta}
\end{bmatrix}
\] respectively.

To represent (9.1.1) as a state space model, we define the state vector as the random walk, that is

\[
\begin{bmatrix}
\alpha_{1t} \\
\alpha_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\alpha_{1,t-1} \\
\alpha_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix}
\]

Let

\[
\begin{bmatrix}
\varepsilon_{1t}^* \\
\varepsilon_{2t}^* \\
\varepsilon_{3t}^* \\
\varepsilon_{4t}^*
\end{bmatrix}
\]

be a zero mean white noise process with covariance matrix \(I_4\). We will use the shorthand notation \(\varepsilon_t^*\).

The covariance matrix \[
\begin{bmatrix}
\sigma_{11, \eta} & \sigma_{12, \eta} \\
\sigma_{21, \eta} & \sigma_{22, \eta}
\end{bmatrix}
\] is positive semidefinite (or definite),

it has a square root \[
\begin{bmatrix}
q_{11, \eta} & q_{12, \eta} \\
q_{21, \eta} & q_{22, \eta}
\end{bmatrix}
\].

Also define \[
\begin{bmatrix}
p_{11, \varepsilon} & p_{12, \varepsilon} \\
p_{21, \varepsilon} & p_{22, \varepsilon}
\end{bmatrix} = \begin{bmatrix}
\sigma_{11, \varepsilon} & \sigma_{12, \varepsilon} \\
\sigma_{21, \varepsilon} & \sigma_{22, \varepsilon}
\end{bmatrix}^{\frac{1}{2}}
\]
Then
\[
\begin{bmatrix}
\eta_{1t} \\
\eta_{2t}
\end{bmatrix} =
\begin{bmatrix}
q_{11,\eta} & q_{12,\eta} & 0 & 0 \\
q_{21,\eta} & q_{22,\eta} & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^* \\
\epsilon_{3t}^* \\
\epsilon_{4t}^*
\end{bmatrix}
\]

Similarly, we write
\[
\begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix} =
\begin{bmatrix}
0 & 0 & p_{11,\epsilon} & p_{12,\epsilon} \\
0 & 0 & p_{21,\epsilon} & p_{22,\epsilon}
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^* \\
\epsilon_{3t}^* \\
\epsilon_{4t}^*
\end{bmatrix}
\]

Then
\[
\begin{bmatrix}
\alpha_{1t} \\
\alpha_{2t}
\end{bmatrix} =
\begin{bmatrix}
\alpha_{1,t-1} \\
\alpha_{2,t-1}
\end{bmatrix} +
\begin{bmatrix}
q_{11,\eta} & q_{12,\eta} & 0 & 0 \\
q_{21,\eta} & q_{22,\eta} & 0 & 0
\end{bmatrix}\begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^* \\
\epsilon_{3t}^* \\
\epsilon_{4t}^*
\end{bmatrix}
\]

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} =
\begin{bmatrix}
\alpha_{1t} \\
\alpha_{2t}
\end{bmatrix} +
\begin{bmatrix}
0 & 0 & p_{11,\epsilon} & p_{12,\epsilon} \\
0 & 0 & p_{21,\epsilon} & p_{22,\epsilon}
\end{bmatrix}\begin{bmatrix}
\epsilon_{1t}^* \\
\epsilon_{2t}^* \\
\epsilon_{3t}^* \\
\epsilon_{4t}^*
\end{bmatrix}
\]

where

\[W = H = I_2;\]

\[B =
\begin{bmatrix}
q_{11,\eta} & q_{12,\eta} & 0 & 0 \\
q_{21,\eta} & q_{22,\eta} & 0 & 0
\end{bmatrix}\]

and

\[G =
\begin{bmatrix}
0 & 0 & p_{11,\epsilon} & p_{12,\epsilon} \\
0 & 0 & p_{21,\epsilon} & p_{22,\epsilon}
\end{bmatrix}\]

Given the estimate \(\alpha_{1}(t|t)\) of \(\alpha_{1t}\) the point forecast of \(\alpha_{1,t+\ell}\) is

\[
\begin{bmatrix}
\alpha_{1}(t+\ell|t) \\
\alpha_{2}(t+\ell|t)
\end{bmatrix} =
\begin{bmatrix}
w_{11} & w_{12} \\
w_{21} & w_{22}
\end{bmatrix} \begin{bmatrix}
\alpha_{1}(t|t) \\
\alpha_{2}(t|t)
\end{bmatrix}
\]

(9.1.2)
and the point forecast of \( \begin{bmatrix} x_{1,t+\ell} \\ x_{2,t+\ell} \end{bmatrix} \) is
\[
\begin{bmatrix} x_{1,t+\ell} \\ x_{2,t+\ell} \end{bmatrix}
\]

\[
\begin{bmatrix} x_{1}(t+\ell|t) \\ x_{2}(t+\ell|t) \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} \alpha_{1}(t+\ell|t) \\ \alpha_{2}(t+\ell|t) \end{bmatrix} \quad (9.1.3)
\]

But

\[
H = W = I_{2}
\]

so that

\[
\begin{bmatrix} x_{1}(t+\ell|t) \\ x_{2}(t+\ell|t) \end{bmatrix} = \begin{bmatrix} \alpha_{1}(t+\ell|t) \\ \alpha_{2}(t+\ell|t) \end{bmatrix}
\]

\[
= \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^{T} \begin{bmatrix} \alpha_{1}(t|t) \\ \alpha_{2}(t|t) \end{bmatrix}
\]

\[
= \begin{bmatrix} \alpha_{1}(t|t) \\ \alpha_{2}(t|t) \end{bmatrix} : W^{T} = I_{2} = I_{2}
\]

That is, given the estimate

\[
\begin{bmatrix} \alpha_{1}(t|t) \\ \alpha_{2}(t|t) \end{bmatrix} = \begin{bmatrix} \mu_{1}(t|t) \\ \mu_{2}(t|t) \end{bmatrix}
\]

then the point forecast of \( \begin{bmatrix} x_{1,t+\ell} \\ x_{2,t+\ell} \end{bmatrix} \) is
\[
\begin{bmatrix} x_{1}(t+\ell|t) \\ x_{2}(t+\ell|t) \end{bmatrix} = \begin{bmatrix} \mu_{1}(t|t) \\ \mu_{2}(t|t) \end{bmatrix} \quad (9.1.4)
\]
For MSE matrices we make use of

\[
BB^t = \begin{bmatrix}
\sigma_{11,\eta} & \sigma_{12,\eta} \\
\sigma_{21,\eta} & \sigma_{22,\eta}
\end{bmatrix}
\]

and

\[
GG^t = \begin{bmatrix}
\sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\
\sigma_{21,\varepsilon} & \sigma_{22,\varepsilon}
\end{bmatrix}
\]

Let

\[
\text{Cov} \begin{bmatrix}
\alpha_{1t} - \alpha_1(t|t) \\
\alpha_{2t} - \alpha_2(t|t)
\end{bmatrix} = \begin{bmatrix}
\sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\
\sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)}
\end{bmatrix}
\]

be given, then MSE matrices of state vector forecast errors are:

\( \ell = 1 \)

\[
\begin{bmatrix}
\sigma_{11,\alpha(t+1|t)} & \sigma_{12,\alpha(t+1|t)} \\
\sigma_{21,\alpha(t+1|t)} & \sigma_{22,\alpha(t+1|t)}
\end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \end{bmatrix} \begin{bmatrix}
\sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\
\sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)}
\end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\
w_{12} & w_{22} \end{bmatrix}
\]

\[+ BB^t \]

\[= \begin{bmatrix}
\sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\
\sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)}
\end{bmatrix} + BB^t \quad \text{because} \quad W = I_2 \]

\( \ell = 2 \)

\[
\begin{bmatrix}
\sigma_{11,\alpha(t+2|t)} & \sigma_{12,\alpha(t+2|t)} \\
\sigma_{21,\alpha(t+2|t)} & \sigma_{22,\alpha(t+2|t)}
\end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \end{bmatrix} \begin{bmatrix}
\sigma_{11,\alpha(t+1|t)} & \sigma_{12,\alpha(t+1|t)} \\
\sigma_{21,\alpha(t+1|t)} & \sigma_{22,\alpha(t+1|t)}
\end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\
w_{12} & w_{22} \end{bmatrix}
\]

\[+ BB^t \]
\[ \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\ \sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} BB^t \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix} + BB^t \]

\[ \begin{bmatrix} \sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\ \sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)} \end{bmatrix} + 2BB^t \quad \text{(because } W = I_2\text{)} \]

\( \ell \in \mathbb{N}, \, \ell \geq 3 \)

We use induction. Suppose that

\[ \begin{bmatrix} \sigma_{11,\alpha(t+\ell-1|t)} & \sigma_{12,\alpha(t+\ell-1|t)} \\ \sigma_{21,\alpha(t+\ell-1|t)} & \sigma_{22,\alpha(t+\ell-1|t)} \end{bmatrix} \]

\[ = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\ \sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}^{\ell-1} + BB^t + \sum_{j=1}^{\ell-2} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^{j} BB^t \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}^{j} \]

then

\[ \begin{bmatrix} \sigma_{11,\alpha(t+\ell|t)} & \sigma_{12,\alpha(t+\ell|t)} \\ \sigma_{21,\alpha(t+\ell|t)} & \sigma_{22,\alpha(t+\ell|t)} \end{bmatrix} \]

\[ = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,\alpha(t+\ell-1|t)} & \sigma_{12,\alpha(t+\ell-1|t)} \\ \sigma_{21,\alpha(t+\ell-1|t)} & \sigma_{22,\alpha(t+\ell-1|t)} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix} + BB^t \]

\[ = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} \sigma_{11,\alpha(t|t)} & \sigma_{12,\alpha(t|t)} \\ \sigma_{21,\alpha(t|t)} & \sigma_{22,\alpha(t|t)} \end{bmatrix} \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}^{\ell} + BB^t \]

\[ + \sum_{j=1}^{\ell-1} \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}^{j} BB^t \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix}^{j} \quad \text{(using induction step } t+\ell-1)\]
For each \( \ell \in \mathbb{N} \), the MSE forecast matrix is

\[
\begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix} + \ell BB' + \sum_{j=1}^{\ell-1} \Sigma BB'
\]

(because \( W = I_2 \))

\[
= \begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix} + \ell BB'
\]

\[
= \begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix} + \ell \begin{bmatrix}
\sigma_{11,\eta} & \sigma_{12,\eta} \\
\sigma_{21,\eta} & \sigma_{22,\eta}
\end{bmatrix}
\]

\[(9.1.5)\]

For the random walk plus noise, if given

\[
\begin{bmatrix}
h_{11} & h_{12} \\
h_{21} & h_{22}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\sigma_{11,\alpha}(t+\ell|t) & \sigma_{12,\alpha}(t+\ell|t) \\
\sigma_{21,\alpha}(t+\ell|t) & \sigma_{22,\alpha}(t+\ell|t)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
h_{11} & h_{12} \\
h_{21} & h_{22}
\end{bmatrix}
\begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix}
\begin{bmatrix}
h_{11} & h_{12} \\
h_{21} & h_{22}
\end{bmatrix} + GG'
\]

\[(9.1.6)\]

For the random walk plus noise, if given

\[
\begin{bmatrix}
\sigma_{11,\alpha}(t|t) & \sigma_{12,\alpha}(t|t) \\
\sigma_{21,\alpha}(t|t) & \sigma_{22,\alpha}(t|t)
\end{bmatrix}
\]

where

\[
\sigma_{ii,\alpha}(t|t) = \text{var}(\mu_i(t|t) - \mu_i(t|t)), \quad i = 1, 2
\]

\[
\sigma_{ij,\alpha}(t|t) = \text{cov}(\mu_i(t|t) - \mu_i(t|t), \mu_j(t|t) - \mu_j(t|t)), \quad i \neq j, \quad i,j = 1, 2
\]

then

\[
\begin{bmatrix}
h_{11} & h_{12} \\
h_{21} & h_{22}
\end{bmatrix} = \begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

and (9.1.6) becomes

...
9.1.2 Bivariate local linear trend
The bivariate local linear trend process is

\[
\begin{bmatrix}
\sigma_{11}(t+t) & \sigma_{12}(t+t) \\
\sigma_{21}(t+t) & \sigma_{22}(t+t)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\sigma_{11}(t+t) & \sigma_{12}(t+t) \\
\sigma_{21}(t+t) & \sigma_{22}(t+t)
\end{bmatrix} + GG' + \ell \begin{bmatrix}
\sigma_{11,\eta} & \sigma_{12,\eta} \\
\sigma_{21,\eta} & \sigma_{22,\eta}
\end{bmatrix} + \begin{bmatrix}
\sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\
\sigma_{21,\varepsilon} & \sigma_{22,\varepsilon}
\end{bmatrix}
\]

(9.1.7)

\[\begin{bmatrix}
x_{1t} \\
x_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t}
\end{bmatrix}
\]

(9.1.8a)

where

\[\begin{bmatrix}
\mu_{1t} \\
\mu_{2t}
\end{bmatrix} = \begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
\]

(9.1.8b)

\[\begin{bmatrix}
\beta_{1t} \\
\beta_{2t}
\end{bmatrix} = \begin{bmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix} + \begin{bmatrix}
w_{1t} \\
w_{2t}
\end{bmatrix}
\]

(9.1.8c)

where \(\varepsilon_{1t}\), \(\varepsilon_{2t}\), \(v_{1t}\), and \(w_{1t}\) are uncorrelated white noise processes with respective (positive definite) covariance matrices given by

\[
\begin{bmatrix}
\sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\
\sigma_{21,\varepsilon} & \sigma_{22,\varepsilon}
\end{bmatrix} = \begin{bmatrix}
p_{11,\varepsilon} & p_{12,\varepsilon} \\
p_{21,\varepsilon} & p_{22,\varepsilon}
\end{bmatrix},
\]

\[
\begin{bmatrix}
p_{11,\varepsilon} & p_{12,\varepsilon} \\
p_{21,\varepsilon} & p_{22,\varepsilon}
\end{bmatrix} = \begin{bmatrix}
p_{11,\varepsilon} & p_{12,\varepsilon} \\
p_{21,\varepsilon} & p_{22,\varepsilon}
\end{bmatrix}.
\]
We consider bivariate white noise process \((e_{1t}^{(1)}, e_{1t}^{(2)}, e_{1t}^{(3)})\) and \((e_{2t}^{(1)}, e_{2t}^{(2)}, e_{2t}^{(3)})\) which are pairwise uncorrelated, each with covariance matrix \(I_2\).

Now

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t} \\
\vdots \\
\beta_{1t} \\
\beta_{2t}
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1} \\
\vdots \\
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix}
\]

\[
+ 
\begin{bmatrix}
0 & 0 & q_{11,v} & q_{12,v} \\
0 & 0 & q_{21,v} & q_{22,v} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\epsilon_{1t}^{(1)} \\
\epsilon_{1t}^{(2)} \\
\vdots \\
\epsilon_{1t}^{(3)}
\end{bmatrix}
+ 
\begin{bmatrix}
b_{11,w} & b_{12,w} \\
b_{21,w} & b_{22,w}
\end{bmatrix}
\begin{bmatrix}
\epsilon_{2t}^{(1)} \\
\epsilon_{2t}^{(2)} \\
\epsilon_{2t}^{(3)}
\end{bmatrix}
\]

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t} \\
\vdots \\
\beta_{1t} \\
\beta_{2t}
\end{bmatrix}
+ 
\begin{bmatrix}
p_{11,c} & p_{12,c} \\
p_{21,c} & p_{22,c}
\end{bmatrix}
\begin{bmatrix}
\epsilon_{t}^{(1)} \\
\epsilon_{t}^{(2)} \\
\epsilon_{t}^{(3)}
\end{bmatrix}
\]

We use

\[
BB' = \begin{bmatrix}
\sigma_{11,v} & \sigma_{12,v} \\
\sigma_{21,v} & \sigma_{22,v}
\end{bmatrix} + 
\begin{bmatrix}
\sigma_{11,w} & \sigma_{12,w} \\
\sigma_{21,w} & \sigma_{22,w}
\end{bmatrix}
\]
We define

$$W^0 = I_4$$

Now

$$W = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$W^2 = WW = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We assume that for some $j > 2, j \in \mathbb{N}$,

$$W^{j-1} = \begin{bmatrix} 1 & 0 & j-1 & 0 \\ 0 & 1 & 0 & j-1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

then

$$W^j = W^{j-1}W = WW^{j-1} = \begin{bmatrix} 1 & 0 & j & 0 \\ 0 & 1 & 0 & j \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
Thus, combining with $W^0$ we have defined,

$$W^j = \begin{bmatrix} 1 & 0 & \cdots & j & 0 \\ 0 & 1 & \cdots & 0 & j \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \quad j = 0, 1, 2, \ldots$$

$$H = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \end{bmatrix}$$

$$HW^j = \begin{bmatrix} 1 & 0 & \cdots & j & 0 \\ 0 & 1 & \cdots & 0 & j \end{bmatrix}, \quad j = 0, 1, 2, \ldots$$

Given the estimate

$$\begin{bmatrix} \mu_1(t|t) \\ \mu_2(t|t) \\ \beta_1(t|t) \\ \beta_2(t|t) \end{bmatrix}$$

the point forecast is

$$\begin{bmatrix} x_1(t+\ell|t) \\ x_2(t+\ell|t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1(t|t) \\ \mu_2(t|t) \\ \beta_1(t|t) \\ \beta_2(t|t) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \ell & 0 \\ 0 & 1 & 0 & \ell \end{bmatrix} \begin{bmatrix} \mu_1(t|t) \\ \mu_2(t|t) \\ \beta_1(t|t) \\ \beta_2(t|t) \end{bmatrix}$$
\[
\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix} + \ell \begin{bmatrix}
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}, \quad \ell = 1, 2, \ldots \quad (9.1.9)
\]

The above equation still allows \( \ell = 0 \), but then we mean that the estimate
\[
\begin{bmatrix}
x_1(t|t) \\
x_2(t|t)
\end{bmatrix} = \begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix},
\]
which is the level.

From (9.1.9) the point forecasts for the (bivariate) local linear trend is increased by the slope at every next horizon.

Let the covariance matrix
\[
\begin{bmatrix}
\Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t)
\end{bmatrix}
\]
be given, where
\[
\Sigma_{\mu\mu}(t|t) = \text{cov}\left[\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix}\right],
\]
\[
\Sigma_{\beta\beta}(t|t) = \text{cov}\left[\begin{bmatrix}
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}\right],
\]
\[
\Sigma_{\mu\beta}(t|t) = \text{cov}\left[\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t)
\end{bmatrix}, \begin{bmatrix}
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}\right]
\]
and
\[
\Sigma_{\beta\mu}(t|t) = \Sigma_{\mu\beta}(t|t).
\]

For convenience we will write
\[
HW^j = \begin{bmatrix}
I_2 & jI_2
\end{bmatrix}, \quad j = 0, 1, \ldots
\]
then
\[ \text{HW}^j_\mathcal{B}(\text{HW}^j_\mathcal{B})' = \Sigma_{vv} + j^2 \Sigma_{ww} \]

\[
= \begin{bmatrix}
\sigma_{11,v} & \sigma_{12,v} \\
\sigma_{21,v} & \sigma_{22,v}
\end{bmatrix} + j^2 \begin{bmatrix}
\sigma_{11,w} & \sigma_{12,w} \\
\sigma_{21,w} & \sigma_{22,w}
\end{bmatrix}
\]

The MSE forecast matrices are given by

\[
\Sigma_{xx}(t+\ell\mid t) = \text{HW}^\ell \Sigma_{xx}(t\mid t)(\text{HW}^\ell)' + \sum_{j=0}^{\ell-1} \text{HW}^j_\mathcal{B}(\text{HW}^j_\mathcal{B})' + GG' \quad (9.1.10)
\]

\( \ell = 1 \)

\[
\begin{bmatrix}
\sigma_{11,x}(t+1\mid t) & \sigma_{12,x}(t+1\mid t) \\
\sigma_{21,x}(t+1\mid t) & \sigma_{22,x}(t+1\mid t)
\end{bmatrix} = \begin{bmatrix} I_2 & I_2 \end{bmatrix} \begin{bmatrix}
\Sigma_{\mu\mu}(t\mid t) & \Sigma_{\mu\beta}(t\mid t) \\
\Sigma_{\beta\mu}(t\mid t) & \Sigma_{\beta\beta}(t\mid t)
\end{bmatrix} \begin{bmatrix} I_2 \\
I_2
\end{bmatrix}
\]

\[+ \begin{bmatrix}
\sigma_{11,v} & \sigma_{12,v} \\
\sigma_{21,v} & \sigma_{22,v}
\end{bmatrix} + \begin{bmatrix}
\sigma_{11,\epsilon} & \sigma_{12,\epsilon} \\
\sigma_{21,\epsilon} & \sigma_{22,\epsilon}
\end{bmatrix}
\]

\[= \Sigma_{\mu\mu}(t\mid t) + \Sigma_{\mu\beta}(t\mid t) + \Sigma_{\beta\mu}(t\mid t) + \Sigma_{\beta\beta}(t\mid t) + \Sigma_{vv} + \Sigma_{ee} \quad (9.1.11a)
\]

\( \ell = 2 \)

\[
\begin{bmatrix}
\sigma_{11,x}(t+2\mid t) & \sigma_{12,x}(t+2\mid t) \\
\sigma_{21,x}(t+2\mid t) & \sigma_{22,x}(t+2\mid t)
\end{bmatrix} = \begin{bmatrix} I_2 & 2I_2 \end{bmatrix} \begin{bmatrix}
\Sigma_{\mu\mu}(t\mid t) & \Sigma_{\mu\beta}(t\mid t) \\
\Sigma_{\beta\mu}(t\mid t) & \Sigma_{\beta\beta}(t\mid t)
\end{bmatrix} \begin{bmatrix} I_2 \\
2I_2
\end{bmatrix}
\]

\[+ \begin{bmatrix}
\sigma_{11,v} & \sigma_{12,v} \\
\sigma_{21,v} & \sigma_{22,v}
\end{bmatrix} + \begin{bmatrix}
\sigma_{11,v} & \sigma_{12,v} \\
\sigma_{21,v} & \sigma_{22,v}
\end{bmatrix} + \begin{bmatrix}
\sigma_{11,w} & \sigma_{12,w} \\
\sigma_{21,w} & \sigma_{22,w}
\end{bmatrix} + \begin{bmatrix}
\sigma_{11,\epsilon} & \sigma_{12,\epsilon} \\
\sigma_{21,\epsilon} & \sigma_{22,\epsilon}
\end{bmatrix}
\]

\[= \Sigma_{\mu\mu}(t\mid t) + 2\Sigma_{\mu\beta}(t\mid t) + 2\Sigma_{\beta\mu}(t\mid t) + 4\Sigma_{\beta\beta}(t\mid t) + 2\Sigma_{vv} + \Sigma_{ww} + \Sigma_{ee} \quad (9.1.11b)
\]
\( \ell = 3 \)

\[
\begin{bmatrix}
\sigma_{11,\sigma}(t+3|t) & \sigma_{12,\sigma}(t+3|t) \\
\sigma_{21,\sigma}(t+3|t) & \sigma_{22,\sigma}(t+3|t)
\end{bmatrix} = \begin{bmatrix} I_2 & 3I_2 \end{bmatrix} \begin{bmatrix} \Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t) \end{bmatrix} \begin{bmatrix} I_2 \\
3I_2
\end{bmatrix}
\]

\[+ \Sigma_{\nu\nu} + (\Sigma_{\nu\nu} + 1^2 \Sigma_{\sigma\sigma}) + (\Sigma_{\nu\nu} + 2^2 \Sigma_{\sigma\sigma}) + \Sigma_{\epsilon\epsilon}\]

\[= \Sigma_{\mu\mu}(t|t) + 3\Sigma_{\mu\beta}(t|t) + 3\Sigma_{\beta\mu}(t|t) + 9\Sigma_{\beta\beta}(t|t) + 3\Sigma_{\nu\nu} + 5\Sigma_{\sigma\sigma} + \Sigma_{\epsilon\epsilon}
\]

(9.1.11c)

For \( \ell \in \mathbb{N}, \ell \geq 4 \), using formula (9.1.10) and others just before it:

\[
\begin{bmatrix}
\sigma_{11,\sigma}(t+\ell|t) & \sigma_{12,\sigma}(t+\ell|t) \\
\sigma_{21,\sigma}(t+\ell|t) & \sigma_{22,\sigma}(t+\ell|t)
\end{bmatrix} = \begin{bmatrix} I_2 & \ell I_2 \end{bmatrix} \begin{bmatrix} \Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t) \end{bmatrix} \begin{bmatrix} I_2 \\
\ell I_2
\end{bmatrix}
\]

\[+ \sum_{j=0}^{\ell-1} (\Sigma_{\nu\nu} + j^2 \Sigma_{\sigma\sigma}) + \Sigma_{\epsilon\epsilon}\]

\[= \Sigma_{\mu\mu}(t|t) + \ell\Sigma_{\mu\beta}(t|t) + \ell\Sigma_{\beta\mu}(t|t) + \ell^2\Sigma_{\beta\beta}(t|t) + \ell\Sigma_{\nu\nu}
\]

\[+ \frac{\ell(\ell-1)(2\ell-1)}{6} \Sigma_{\sigma\sigma} + \Sigma_{\epsilon\epsilon}
\]

by using the formula

\[
\sum_{j=0}^{\ell-1} j^2 = \sum_{j=1}^{\ell-1} j^2
\]

\[= \frac{(\ell-1)\ell(2\ell-1)+1}{6}
\]

\[= \frac{\ell(\ell-1)(2\ell-1)}{6}
\]

By writing in long-hand with
\[ \Sigma_{\mu\beta}(t|t) = \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix} \]

where

\[ \sigma_{ij}(t|t) = \text{cov}\left[\mu_{it} - \mu_1(t|t), \beta_{jt} - \beta_j(t|t)\right], \quad i = 1, 2, \]

then

\[
\begin{bmatrix} \sigma_{11}(t+\ell|t) & \sigma_{12}(t+\ell|t) \\ \sigma_{21}(t+\ell|t) & \sigma_{22}(t+\ell|t) \end{bmatrix} = \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix} \\
+ \ell \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix} + \ell^2 \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix} + \ell^2 \begin{bmatrix} \sigma_{11}(t|t) & \sigma_{12}(t|t) \\ \sigma_{21}(t|t) & \sigma_{22}(t|t) \end{bmatrix}
\]

\[ \begin{bmatrix} \sigma_{11,w} & \sigma_{12,w} \\ \sigma_{21,w} & \sigma_{22,w} \end{bmatrix} + \begin{bmatrix} \sigma_{11,\varepsilon} & \sigma_{12,\varepsilon} \\ \sigma_{21,\varepsilon} & \sigma_{22,\varepsilon} \end{bmatrix} \quad \text{(9.1.11c)} \]

9.1.3 Bivariate damped trend model


The model in question is given by

\[
\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad \text{(9.1.12a)}
\]

where

\[
\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \quad \text{(9.1.12b)}
\]

\[
\begin{bmatrix} \beta_{1t} \\ \beta_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix} \quad \text{(9.1.12c)}
\]

where
and all the other components are (defined exactly) as in the bivariate local linear trend process, (9.1.8).

**Definition (Damped Matrix)**
The matrix \( \Phi \) satisfying property (9.1.13) is called a **damped matrix**.

The SSR is set up as follows:

\[
\begin{bmatrix}
\mu_{1t} \\
\mu_{2t} \\
\vdots \\
\beta_{1t} \\
\beta_{2t}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & \phi_{11} & \phi_{12} \\
0 & 1 & \phi_{21} & \phi_{22} \\
\vdots \\
0 & 0 & \phi_{11} & \phi_{12} \\
0 & 0 & \phi_{21} & \phi_{22}
\end{bmatrix}\begin{bmatrix}
\mu_{1,t-1} \\
\mu_{2,t-1} \\
\vdots \\
\beta_{1,t-1} \\
\beta_{2,t-1}
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
0 & 0 & q_{11,v} & q_{12,v} \\
0 & 0 & q_{21,v} & q_{22,v} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\beta_{1t} \\
\beta_{2t}
\end{bmatrix}
\]

where

\[
W = \begin{bmatrix}
1 & 0 & \phi_{11} & \phi_{12} \\
0 & 1 & \phi_{21} & \phi_{22} \\
0 & 0 & \phi_{11} & \phi_{12} \\
0 & 0 & \phi_{21} & \phi_{22}
\end{bmatrix}
\]
and all the other components are as in model (9.1.8) and its SSR.

\[
W^2 = \begin{bmatrix}
I_2 & \Phi \\
O_2 & \Phi
\end{bmatrix}
\begin{bmatrix}
I_2 & \Phi \\
O_2 & \Phi
\end{bmatrix}
= \begin{bmatrix}
I_2 & \Phi + \Phi^2 \\
O_2 & \Phi^2
\end{bmatrix}
\]

\[
W^3 = W^2W
= \begin{bmatrix}
I_2 & \Phi + \Phi^2 \\
O_2 & \Phi^2
\end{bmatrix}
\begin{bmatrix}
I_2 & \Phi \\
O_2 & \Phi
\end{bmatrix}
= \begin{bmatrix}
I_2 & \Phi + \Phi^2 + \Phi^3 \\
O_2 & \Phi^3
\end{bmatrix}
\]

Assume that the induction step is

\[
W^{j-1} = \begin{bmatrix}
I_2 & \sum_{k=1}^{j-1} \Phi^k \\
O_2 & \Phi^{j-1}
\end{bmatrix}
\text{ for some } j \in \mathbb{N}, j \geq 4
\]

then

\[
W^j = W^{j-1}W
= \begin{bmatrix}
I_2 & \sum_{k=1}^{j-1} \Phi^k \\
O_2 & \Phi^{j-1}
\end{bmatrix}
\begin{bmatrix}
I_2 & \Phi \\
O_2 & \Phi
\end{bmatrix}
= \begin{bmatrix}
I_2 & \Phi + \Phi \sum_{k=1}^{j-1} \Phi^k \\
O_2 & \Phi^j
\end{bmatrix}
\]
\[
\begin{bmatrix}
I_2 & \Phi_+ \frac{j}{2} \Phi^k \\
0 & \Phi_j
\end{bmatrix}
\]
\[
\begin{bmatrix}
I_2 & \sum_{k=1}^{j} \Phi^k \\
0 & \Phi_j
\end{bmatrix}
\]
for \( j \in \mathbb{N}, \ j \geq 4 \)

By defining
\[
\Phi^0 = I_2
\]
and
\[
W^0 = I_4
\]
then
\[
W^j = \begin{bmatrix}
I_2 & \sum_{k=1}^{j} \Phi^k \\
0 & \Phi_j
\end{bmatrix}, \ j = 1, 2, ...
\]
\[
HW^j = \begin{bmatrix}
I_2 & 0
\end{bmatrix} W^j
\]
\[
= \begin{bmatrix}
I_2 & \sum_{k=1}^{j} \Phi^k \\
0 & \Phi_j
\end{bmatrix}, \ j = 1, 2, ...
\]
\[
= \begin{bmatrix}
I_2 & \Phi(I-I_\Phi)^{-1}(I-I_\Phi^j)
\end{bmatrix}
\]
and
\[
HW^0 = \begin{bmatrix}
I_2 & 0
\end{bmatrix}.
\]

Also, using block notation,
\[
(HW^0B)(HW^0B)' = \Sigma_{vv}
\]
\[
(HW^jB)(HW^jB)' = \begin{bmatrix}
I_2 & \sum_{k=1}^{j} \Phi^k \\
0 & \sum_{k=1}^{j} \Phi^k
\end{bmatrix}
\begin{bmatrix}
O_2 & \frac{1}{2} \Sigma_{vv} & O_2 \\
O_2 & O & \frac{1}{2} \Sigma_{ww}
\end{bmatrix}
\begin{bmatrix}
O_2 & O_2 \\
O_2 & \sum_{k=1}^{j} \Phi^k
\end{bmatrix}
\begin{bmatrix}
I_2 \\
O_2 & \Sigma_{ww}
\end{bmatrix}
The $\ell$-step ahead forecast, where we are given

$$
\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t) \\
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}
$$

is given by

$$
\begin{bmatrix}
x_1(t+\ell|t) \\
x_2(t+\ell|t)
\end{bmatrix} = HW^\ell
\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t) \\
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}
$$

$$
= \begin{bmatrix}
I_2 \\
\Phi(I-\Phi)^{-1}(I-\Phi^\ell)
\end{bmatrix}
\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t) \\
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\mu_1(t|t) + \phi_{11} \phi_{12} [1-\phi_{11} -\phi_{12}]^{-1} \\
\mu_2(t|t) + \phi_{21} \phi_{22} [1-\phi_{21} -\phi_{22}]^{-1}
\end{bmatrix}
\begin{bmatrix}
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix}
$$

(9.1.14a)

for large $\ell$ as (9.1.13) would suggest.
For forecast MSE matrix, we use block notation on (9.1.10), suppose

\[
\begin{bmatrix}
\Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t)
\end{bmatrix}
\]
is given. Then:

\[
\ell = 1
\]

\[
\begin{bmatrix}
\sigma_{11,\mathbf{x}}(t+1|t) & \sigma_{12,\mathbf{x}}(t+1|t) \\
\sigma_{21,\mathbf{x}}(t+1|t) & \sigma_{22,\mathbf{x}}(t+1|t)
\end{bmatrix} = \begin{bmatrix} I_2 & \Phi \end{bmatrix} \begin{bmatrix}
\Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t)
\end{bmatrix} \begin{bmatrix} I_2 \\
\Phi'
\end{bmatrix}
\]

\[+
\Sigma_{vv} + \Sigma_{ee}
\]

\[
= \Sigma_{\mu\mu}(t|t) + \Phi \Sigma_{\beta\mu}(t|t) + \Sigma_{\beta\mu}(t|t) \Phi' + \Phi \Sigma_{\beta\beta}(t|t) \Phi' + \Sigma_{vv} + \Sigma_{ee}
\]

(9.1.15a)

\[
\ell = 2
\]

\[
\begin{bmatrix}
\sigma_{11,\mathbf{x}}(t+2|t) & \sigma_{12,\mathbf{x}}(t+2|t) \\
\sigma_{21,\mathbf{x}}(t+2|t) & \sigma_{22,\mathbf{x}}(t+2|t)
\end{bmatrix} = \begin{bmatrix} I_2 & \Phi \Phi^2 \\
\end{bmatrix} \begin{bmatrix}
\Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t)
\end{bmatrix} \begin{bmatrix} I_2 \\
\Phi' + (\Phi')^2
\end{bmatrix}
\]

\[+
\Sigma_{vv} + \left\{ \Sigma_{vv} + \Phi \Sigma_{ww} \Phi' \right\} + \Sigma_{ee}
\]

(9.1.15b)

For \( \ell \in \mathbb{N}, \ell \geq 3 \)

\[
\begin{bmatrix}
\sigma_{11,\mathbf{x}}(t+\ell|t) & \sigma_{12,\mathbf{x}}(t+\ell|t) \\
\sigma_{21,\mathbf{x}}(t+\ell|t) & \sigma_{22,\mathbf{x}}(t+\ell|t)
\end{bmatrix} = \begin{bmatrix} I_2 & \Sigma \Phi^k \\
\end{bmatrix} \begin{bmatrix}
\Sigma_{\mu\mu}(t|t) & \Sigma_{\mu\beta}(t|t) \\
\Sigma_{\beta\mu}(t|t) & \Sigma_{\beta\beta}(t|t)
\end{bmatrix} \begin{bmatrix} I_2 \\
\Sigma (\Phi')^k
\end{bmatrix}
\]

\[+
\Sigma_{vv} + \sum_{j=1}^{\ell-1} \Sigma \Phi \Phi^{-1} (I-\Phi) \Sigma_{ww} \left[ I-(\Phi')^j \right] (I-\Phi')^{-1} \Phi' + \Sigma_{ee}
\]
9.1.4 Discussion
The point forecast for the random walk plus noise, seen at (9.1.4), is nonchanging. This is consistent with the assumption that the model explains the (local) level. The nonstationarity of this model is displayed by the forecast MSE (9.1.7) which, apart from being linear on $\Sigma_{\alpha\alpha}(t|t)$ and $\Sigma_{\epsilon\epsilon}$, has the term $\ell \Sigma_{\eta\eta}$ where $\ell = 1, 2, \ldots$. This shows a continuous increase for increasing forecast horizon $\ell$, that is, it is unbounded in the long-run.

The point forecast (9.1.9) for the local linear trend is "increased" by the slope of increasing $\ell$. (We may have one or both $\beta_1(t|t)$ being negative, "increase" simply implies being pooled in the direction given by the signs.) This slope gives us the uncertainty about point forecast because it is unbounded as horizon $\ell$ increases. Also, this forecast already displays nonstationarity which was not displayed by the point forecast of random walk plus noise, (9.1.4). The forecast MSE (9.1.11) diverges even quicker than that of random walk plus noise. The terms which we use in this comparison are

$$
\ell \Sigma_{\mu\mu}(t|t), \ell \Sigma_{\beta\mu}(t|t), \ell \Sigma_{\eta\eta}, \ell^2 \Sigma_{\beta\beta}(t|t) \text{ and } \frac{\ell(\ell-1)(2\ell-1)}{6} \Sigma_{ww}
$$

for the local linear trend model, as compared to just

$$
\ell \Sigma_{\eta\eta}
$$

For the damped trend model, we have a situation 'between random walk plus noise and local linear trend' in that the trend is there in addition to the random walk plus noise, but unlike in the local linear trend, this trend is
limited by the damped matrix which, according to its nature (cf. property (9.1.13)), it damps the trend. As a result, in the long run this trend is not effective. The point forecast at (9.1.14) shows that though it is shifting the forecasts from the given point \( \mu(t|t) \), this shift is not 'too big', is convergent (and is bounded). Moreover, the coefficient in (9.1.14b) of \( \beta(t|t) \) is the limit of the damp. From the forecast MSE, the only unbounded parts are \( \ell \Sigma_{YV} \) (which is clear in (9.1.15) and \( \ell \Phi(I-\Phi)^{-1} \Sigma_{WW}(I-\Phi)^{-1} \Phi^t \) which will come out of the summation parts, all the other parts in the sum have \( \Phi^j \) which therefore will converge. The damped trend model, because of the trend, therefore increases quicker than the random walk plus noise, whereas because the trend is damped, it increases slower than the local linear trend model.

### 9.2 FORECAST REGIONS

Johnson & Wichern (1992: 132)

We recall that a 100(1-\( \alpha \))% forecast region \( \ell \) periods ahead is given by

\[
\begin{bmatrix}
    z_{1t+\ell} - x_{1t}(\ell)
    \\
    z_{2t+\ell} - x_{2t}(\ell)
\end{bmatrix} \Sigma_{XX}^{-1}(t+\ell|t) \begin{bmatrix}
    z_{1t+\ell} - x_{1t}(\ell)
    \\
    z_{2t+\ell} - x_{2t}(\ell)
\end{bmatrix} \leq \chi^2_2(\alpha)
\]  

(9.2.1)

For 90% forecast region, \( \chi^2_2(0.10) = 4.61 \).

The centre for the region is given by \( \begin{bmatrix} x_{1t}(\ell) \\ x_{2t}(\ell) \end{bmatrix} \), and our conclusion from previous sections is that for the random walk plus noise model, this centre will not change, for the trend model it will always move and for the damped trend it will change for some time, and there will be a point beyond which it will not go.

For illustration purposes we assume that in all the processes, the white noise processes all have covariances \( I_2 \), all initial values and initial covariances are:
\[
\begin{bmatrix}
\mu_1(t|t) \\
\mu_2(t|t) \\
\beta_1(t|t) \\
\beta_2(t|t)
\end{bmatrix} = 
\begin{bmatrix} 1 \\
1 \\
1 \\
1 
\end{bmatrix},
\begin{bmatrix}
\Sigma_{\mu}(t|t) : \Sigma_{\beta}(t|t)
\end{bmatrix} = 
\begin{bmatrix}
1.00 & 0 & 0.50 & 0.75 \\
0 & 1.00 & 0 & 0 \\
0.50 & 0 & 1.00 & 0 \\
0.75 & 0 & 0 & 1.00 
\end{bmatrix}
\]

and for the damped matrix \( \Phi \) we choose

\[
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix} = 
\begin{bmatrix}
0.1 & 0 \\
0 & 0.1
\end{bmatrix}
\]

The choice of values is very trivial, simply for convenience and some shapes are as they will be because of these values. For our comparison they are enough, and with \( \ell = 1 \) and \( \ell = 3 \), we are going to deduce the relative behavior of forecasts.

For illustration we plot 90% forecast regions, then \( \chi^2_2(0.10) = 4.61 \). Then \( c = \sqrt{4.61} \).

9.2.1 Random walk plus noise

\( \ell = 1 \)

Using (9.1.4) and (9.1.7) respectively, we obtain

\[
\text{centre } \begin{bmatrix} x_{1t}(1) \\ x_{2t}(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}
\]

\[
\text{MMSE } \begin{bmatrix} \Sigma_x(1) \end{bmatrix} = I_2 + I_2 + I_2
\]

\[
= \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}
\]

The forecast region as defined by (9.2.1) is

\[
\frac{1}{3}[(x-1)^2 + (y-1)^2] < 4.61,
\]
a circle with centre $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and radius $\sqrt{3 \times 4.61} = 3.72$.

$t = 3$

Equations (9.1.4) and (9.1.7) give 
centre $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

and MMSE

$$\Sigma_x(3) = I_2 + 3I_2 + I_2 = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$$

(9.2.1) gives the 90% forecast region as a circle with centre $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and radius $\sqrt{5(4.61)} = 4.80104$.

The centre does not shift but the region (circle) expands as the radius increases. The circle simply displays "close to stationary" of a process as the shape of region is constant. The reason for the circular shape is because of (many) assumptions (such as equal variances and lack of correlation). If we had correlated variables we would be having ellipses, but they would still retain centres (point forecast is constant) and direction of axis would not change.

9.2.2 Bivariate local linear trend

$t = 1$

(9.1.9) and (9.1.11a) give

centre $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$
and MMSE

\[
I_2 + \begin{bmatrix} 0.50 & 0 \\ 0.75 & 0 \end{bmatrix} + \begin{bmatrix} 0.75 & 0.75 \\ 0 & 0 \end{bmatrix} + I_2 + I_2 + I_2
\]

\[
= \begin{bmatrix} 5.00 & 0.75 \\ 0.75 & 4.00 \end{bmatrix} \quad v_1 = \begin{bmatrix} 0.882 \\ 0.472 \end{bmatrix} \quad \theta = 28.15^0
\]

\[
\lambda_1 = 5401, \quad \lambda_2 = 3.599
\]

\[
c_{\lambda_1} = 4.471, \quad c_{\lambda_2} = 4.073
\]

\( \ell = 3 \)

(9.1.9) gives

\[
\text{centre } \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\]

(9.1.11c) the MMSE

\[
\Sigma_{xx}(3) = I_2 + 3\begin{bmatrix} 0.50 & 0 \\ 0.75 & 0 \end{bmatrix} + 9I_2 + 3I_2 + 5I_2 + I_2
\]

\[
= \begin{bmatrix} 22.00 & 2.25 \\ 2.25 & 19.00 \end{bmatrix} \quad v_1 = \begin{bmatrix} 0.882 \\ 0.472 \end{bmatrix} \quad \theta = 28.15^0
\]

\[
\lambda_1 = 23.204, \quad \lambda_2 = 17.796
\]

\[
c_{\lambda_1} = 10.343, \quad c_{\lambda_2} = 9.058
\]

The centre shifts up along the line \( z_{1t} = z_{2t} \), and the expansion of the regions for increasing \( \ell \) is "explosive", seen from the huge increase in length of both axis. This is despite the small values given to the parameters for easy calculations. In general, the rate of unboundedness is very high.
9.2.3 Damped trend model

\( \ell = 1 \)

Using

\[
HW = \begin{bmatrix} 1 & \Phi \\ \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0.1 \end{bmatrix}
\]

\[
\text{centre } \begin{bmatrix} x_{1t}(1) \\ x_{2t}(1) \end{bmatrix} = HW \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 1.1 \end{bmatrix}
\]

and (9.1.15a) gives

\[
\Sigma_{xx}(1) = I + 0.1 \begin{bmatrix} 0.50 & 0 \\ 0.50 & 0.75 \end{bmatrix} + (0.1)I + I + I
\]

\[
= \begin{bmatrix} 3.16 & 0.05 \\ 0.05 & 3.01 \end{bmatrix}
\]

\[
\mathbf{v}_1 = \begin{bmatrix} 0.958 \\ 0.287 \end{bmatrix}, \quad \theta = 16.68^\circ
\]

\[
\lambda_1 = 3.175, \quad \lambda_2 = 2.995
\]

\[
c \sqrt{\lambda_1} = 3.826, \quad c \sqrt{\lambda_2} = 3.716
\]

\( \ell = 3 \)

From (9.1.14a) we need to evaluate the following expression.

\[
\Phi(I-\Phi)^{-1}(I-\Phi^3) = 0.1I(0.9I)^{-1}(0.999I)
\]

\[= 0.111I\]

\[
\text{centre } \begin{bmatrix} x_{1t}(3) \\ x_{2t}(3) \end{bmatrix} = \begin{bmatrix} I & \Phi(I-\Phi)^{-1}(I-\Phi^3) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

\[
\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]
Expression (9.1.15c) gives

\[
\Sigma_{xx}(3) = \Sigma_{\mu}(t|t) + \Phi^* \Sigma_{\beta}(t|t) + \Sigma_{\mu} \Phi^* + \Phi^* \Sigma_{\beta}(t|t) \Phi^* + 3 \Sigma_{vv}
\]

\[
+ \Phi \Sigma_{ww} \Phi' + (\Phi + \Phi^2) \Sigma_{ww} (\Phi' + (\Phi')^2) + \Sigma_{EE}
\]

where

\[
\Phi^* = \Phi + \Phi^2 + \Phi^3.
\]

Therefore

\[
\Sigma_{xx}(3) = I + (0.111) \begin{bmatrix} 0.50 & 0 \\ 0.75 & 0 \end{bmatrix} + (0.111) \begin{bmatrix} 0.50 & 0.75 \\ 0 & 0 \end{bmatrix}
\]

\[
+ (0.111)^2 I + 3 I + (0.01) I + (0.11)^2 I + I
\]

\[
= \begin{bmatrix} 5.201 & 0.083 \\ 0.083 & 5.034 \end{bmatrix} \quad v_1 = \begin{bmatrix} 0.958 \\ 0.287 \end{bmatrix} \quad \theta = 16.68^0
\]

\[
\lambda_1 = 5.226 \quad \lambda_2 = 5.009
\]

\[
c_4 \lambda_1 = 4.908, \quad c_4 \lambda_2 = 4.805
\]

(See Figure 9.1 on graph page 213a.)

The shift in centre is very slow, and in the long run is not meaningful. The size of the region does increase, but not as quick as in local linear trend, but still quicker than for the random walk plus noise. The effect of the trend is slight, but it is minimized by the damping \( \Phi \).

Remark

The unbounded nature of forecast regions signals the difficulty in forecasting accurately several steps ahead.
Figure 9.1: 90% 3-step ahead forecast region for the damped trend model.
CHAPTER 10

COMPARISON OF (BIVARIATE) ARIMA
AND STRUCTURAL MODELS

Harvey (1984: 245)

We compare bivariate ARIMA and bivariate structural models on the basis of
the models discussed in the last eight chapters and the properties considered.
The state space models and the Kalman filter became a "unified framework
for gaining insight in the structures and implications" of these models. The
structure of a model could well be understood by its long term behaviour, and
the forecast behavior of certain models could display this, for example the
random walk plus noise has constant point forecasts for all forecast horizons,
and the damped trend has, in the long run, a forecast whose trend
contribution is damped to almost null. All these are provided conveniently by
the Kalman recursions.

10.1 GENERALIZATIONS OF EXPONENTIAL SMOOTHING BY
CLASSES OF STATISTICAL MODELS

Newbold & Bos (1990: 356)

In the generalization of simple bivariate exponential smoothing algorithm by
ARIMA(0,1,1) (that is IMA(1,1)), we note that the parameters $\Theta$ of
ARIMA and $A$ of exponential smoothing are related by $\Theta = 1 - A$, so
evaluating one is evaluating the other. In Remark 2 of this relationship, we
note that the parameter $\Theta$ can be estimated only when $1 - \Theta$ is invertible
which means when $(1 - \Theta)^{-1}$ exists. Hence the $\Theta$ of exponential
smoothing, if we insist on this relationship, cannot be chosen for the merits of
the exponential smoothing algorithm, but for the sake of the relationship. We
can distort the relationship in this way. All we are saying is that not all
exponential smoothing methods have $\Theta$ satisfying the property.

Harvey (1990: 346-361)

The problem of relationship with parameters for Holt's and damped
exponential algorithms are even worse (cf. Newbold & Box 1990: pp. 356-361),
as they suggest that the smoothing parameters be chosen 'objectively' and not for the merits of exponential smoothing itself. Whenever one is in bivariate situation which is an analogue of univariate case, the problems of univariate are inherited by bivariate in multiples, and also it is logical to note that because the problem with \( \theta = 1-\alpha \) led to a worse problem when considering \( \Theta = I - A \), then these two, plus the problems we have with \( \theta_1 = 2 - \alpha - \alpha\beta \) and \( \theta_2 = -(1 - \alpha) \) lead to an even more difficult problem with \( \Theta_1 = 2I - A_1 - A_2A_1 \) and \( \Theta_2 = -(I - A_1) \) which is for Holt's algorithm and ARIMA(0,2,2) while for ARIMA(1,1,2) we have \( \Theta_1 = I + \Phi - A_1 - \Phi A_2A_1 \) and \( \Theta = -\Phi(I - A_1) \). The forms of relationships are fine but the values of related parameters distort the relationships.

Harvey (1984: 165-6, 171; 1989: 13)

On the other hand the error-correction forms as seen in Harvey (1984) and our work in Chapter 2, (assumption 2.2.1) that the bivariate vector is the sum of level (bivariate) and irregular components), the error-correction form implies a structural model. On the relevant portions where exponential smoothing algorithms were shown to be suggesting structural models, the "transformation" was not complicated and the assumption in different exponential smoothing algorithms were conserved in structural models. Harvey (1989: 13) state that "error-correction mechanism can be employed, and the inclusion of the unobserved time series components does not affect the model selection methodology in any fundamental way". These unobserved time series components are the parameters of structural models.

### 10.2 COINTEGRATION PROPERTIES

Reinsel (1993: 42), Lütkepohl (1991: 423)

Both ARIMA and structural models require a common nonstationary factor with a loading vector (or matrix) such that when premultiplying the loading vector (matrix) by a cointegrating vector (matrix) the whole term containing the common factor is wiped away, that is if \( x_t = \alpha f_t + \text{other terms including error} \), where \( f_t \) is the common factor (which is nonstationary), the product \( \alpha' \theta_f \) for an integrating vector \( \alpha \) must be zero for every common factor \( f_t \). Thus \( \alpha' \theta = 0 \). The matrix analogue is
The structural components do have matrix parameters on rare bases while ARIMA is defined in terms of them (the AR and MA matrices). The structural time series may have it when introducing damp or expansion in matrix form and when emphasizing nonSUTSE, and because not all nonstationary bivariate series have matrices satisfying (10.2.1) lack of cointegration seems more likely to occur in bivariate ARIMA than in bivariate structural models.

Forecast properties of cointegrated models are similar in both ARIMA and structural models, for example as we said about them in first paragraph of this chapter.

10.3 BIVARIATE AUTOREGRESSIONS

Examples used in Chapter 7, Examples 7.1 and 7.2 show that model with autoregression introduced in levels or errors do not handle cointegration properties well, except when a dynamic model is used, that is when autoregression is introduced on the term \( x_t \) itself. The autoregression (AR) may be more associated with ARIMA than with structural model because of AR. In this case it counts against the ARIMA model in the handling of cointegration.

10.4 CONCLUSION

Harvey (1989: 13, 65, 75)

Bivariate ARIMA models are generally not bad as far as forecasting aspects are concerned, but care should be taken when introducing any additional properties such as cointegration on the model. The form of ARIMA when this property is introduced for example, must not be as in additive model or autoregression model of Chapter 7 and 8. The dynamic model of same chapters is a form that can be interpreted as structural by defining \( z_t^* = \Phi(L)x_t \) and proceed with \( z_t^* \). This dynamic model is the one which was found to be "robust" against inefficiency in handling cointegration.

Forecasting using structural models do not indicate problems with the class of models itself even when conditions are added on the model, in fact structural models seem to show property of being "dynamic" (recall the phrase:
components with direct interpretation). Except under causality which we did not look at on structural models, all other conditions imposed on the structural models, "interpretation is direct" as Harvey would say.

Lastly, on the representation of exponential smoothing algorithms, the ARIMA models will not be able to provide values for certain exponential forms, for example where $A^{-1} = (I-\Theta)^{-1}$ does not exist as we said in section 10.1. There will therefore be those bivariate exponential smoothing algorithms whose ARIMA representations are flawed. On the structural models, the error-corrections imply structural models without any visible flaws.

Therefore, for the merits of bivariate exponential smoothing algorithms, the bivariate structural time series models provide better generalizations (than ARIMA) whereas ARIMA representations are not impressive as generalizations of exponential smoothing algorithms.
APPENDIX A

RESULTS USED IN THE DISSERTATION

Theorem A.1

Singular Value Decomposition Theorem (SVD)
Aoki M (1987: 70)
Let \( A \) be a real \( m \times n \) matrix of rank \( r \). Then there exist matrices \( U_1: m \times r \) and \( V_1: n \times r \) with orthonormal columns, that is

\[
U_1^T U_1 = I_r \tag{A.1.1}
\]
\[
V_1^T V_1 = I_r \tag{A.1.2}
\]

and a unique real diagonal matrix \( D = \alpha_1 I_r \) with positive diagonal elements \( \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0 \) such that

\[
A = U_1 D V_1^T \tag{A.1.3}
\]

A more general result when \( r < n \) is the representation (Aoki 1987: 70)

\[
A = U D V^T \tag{A.1.4}
\]

where

\[
UU^T = I_m, \quad VV^T = I_n
\]

\[
U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}^{(m \times r)} \begin{bmatrix} I_r \end{bmatrix}^{(r \times (m-r))}
\]

\[
V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}^{(n \times r)} \begin{bmatrix} I_r \end{bmatrix}^{(r \times (n-r))}
\]
\[
D = \begin{bmatrix}
D_{\alpha} & 0 \\
(r \times r) & (r \times (n-r)) \\
0 & (m-r) \times (n-r)
\end{bmatrix}
\]  
(A.1.5)

If \( r = n \), \( V_1 = V \), \( U_1 = U \) and \( D_\alpha = D \) so that (A.1.3) and (A.1.4) are exactly the same.

Starting from (A.1.4) we show that (A.1.3) is attained, and thereafter we derive (A.1.3) in the proof of Theorem A.1

\[
A = UDV'
\]

That is,

\[
(A.1.4) \implies (A.1.3)
\]

Proof of Theorem A.1

\[
(A' A)' = A'(A')' = A' A
\]

so that \( A' A \) is symmetric, and by Graybill (1983: 18: Theorems 1.7.6/1.7.7) rank \( (A' A) = r \) and \( A' A \) is either positive definite (if \( r = n \)) or positive semidefinite (if \( r < n \)).

By Graybill (1983: 48: Theorem 3.4.4) there exists an orthonormal matrix \( V: n \times n \) and a diagonal matrix \( D_{\lambda} \) with \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0 \) such that

\[
V' A' A V = D_{\lambda}
\]  
(A.1.6)

where \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0 \) and \( \lambda_{r+1} = \ldots = \lambda_n = 0 \)
We call $\lambda_1, \ldots, \lambda_r$ the singular values. By defining $\alpha_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, r$ and $D_\alpha$ a diagonal matrix of the $\alpha_i$'s, $i = 1, \ldots, r$, we write

$$D_\lambda = \begin{bmatrix} D_\alpha^2 & 0 \\ 0 & 0 \end{bmatrix}$$

as in (A.1.5). We note that $D_\alpha$ has size $r \times r$ and $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_r > 0$. Let

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

(A.1.8)

then

$$I_n = V V'$$

$$= \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

$$= V_1 V_1' + V_2 V_2'$$

That is

$$I_n = V_1 V_1' + V_2 V_2'$$

(A.1.9)

By substituting (A.1.7) and (A.1.8) in (A.1.6) and interchanging the left- and right-hand sides we obtain

$$\begin{bmatrix} D_\alpha^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} V_1' (A' A) V_1 \\ V_2' \end{bmatrix}$$

We deduce that

$$V_1' A' AV_1 = D_\alpha^2$$

(A.1.10)

and
The last equation implies that
\[ AV_2 = 0 \]  \hspace{1cm} (A.1.11)

Premultiplying (A.1.9) by \( A \) we have
\[ A = AV_1 V_1' + AV_2 V_2' \]
\[ = AV_1 V_1' \] \hspace{1cm} (from (A.1.11): \( AV_2 = 0 \)).

That is
\[ A = AV_1 V_1' \]  \hspace{1cm} (A.1.12)

We define \( U \) as follows:
\[ U_1 = AV_1 D_{\alpha}^{-1} \]  \hspace{1cm} (A.1.13)

then by postmultiplying (A.1.13) by \( D_{\alpha}' \), then
\[ AV_1 = U_1 D_{\alpha} \]  \hspace{1cm} (A.1.14)

We substitute (A.1.14) in the right-hand side of (A.1.12) to obtain
\[ A = U_1 D_{\alpha} V_1' \]
which is (A.1.3). We need to show also that \( U_1 \) and \( V_1 \) satisfy (A.1.1) and (A.1.2).

Using (A.1.13) we obtain (A.1.1) as follows:
\[ U_1' U_1 = (AV_1 D_{\alpha}^{-1})' AV_1 D_{\alpha}^{-1} \]
\[ = D_{\alpha}^{-1}(V_1' A V_1)D_{\alpha}^{-1} \]
\[ = D_{\alpha}^{-1} \cdot D_{\alpha}^2 \cdot D_{\alpha}^{-1} \] \hspace{1cm} : by (A.1.10)
\[ = I_r \]

By the orthonormality of \( V \), we have:
\[ V'V = I_n \]  \hspace{1cm} (A.1.15)
Using (A.1.8) in (A.1.15):

\[
\begin{bmatrix}
V'_1 \ V'_1 \\

V'_2
\end{bmatrix}
= 
\begin{bmatrix}
I_r & O \\

O & I_{n-r}
\end{bmatrix} 
\]

Multiplying the matrices we have

\[
\begin{bmatrix}
V'_1 \ V'_1 \\

V'_2 \ V'_2
\end{bmatrix}
= 
\begin{bmatrix}
I_r & O \\

O & I_{n-r}
\end{bmatrix}
\]

From what we can deduce from (A.1.16) all we need (for now) is:

\[V'_1V'_1 = I_r\]

which is (A.1.2).

The following corollary to the above theorem is presented.

**Corollary A.2**

Any matrix \(A: m \times n\) of rank \(r \leq \min(m,n)\) may be written as

\[A = HC\]

where \(H: m \times r\) and \(C: r \times n\) are both of rank \(r\).

**Proof (outline)**

One approach is to set (in (A.1.3)/(A.1.4))

\[H = UD_{\alpha}^{1/2}\]

\[C = D_{\alpha}^{1/2}V\]

then

\[A = HC\]

**Theorem A.3 (SDR)**

Let \(A: n \times n\) be a matrix of full rank \(n\) with linearly independent right eigenvectors \(u_1, \ldots, u_n\) and \(n\) eigenvalues \(\lambda_1 > \lambda_2 > \ldots > \lambda_n\). Then, with

\[U = [u_1, u_2, \ldots, u_n],\]
Also, there exists a matrix \( V \) of row vectors \( v_i^T \), \( i = 1, \ldots, n \) such that

\[
A = UDV
\]

where \( D \) is a diagonal matrix with \( \lambda_i \) as entries. This is known as the eigenvalue decomposition of \( A \).

**Proof (outline)**

The \( u_i \) are right eigenvectors with \( \lambda_i \) as eigenvalues, then

\[
Au_i = \lambda_i u_i
\]

\[
\Rightarrow A[u_1, u_2, \ldots, u_n] = [\lambda_1 u_1, \lambda_2 u_2, \ldots, \lambda_n u_n]
\]

or

\[
Au = UD
\]

which proves (A.3.1).

Now, the \( u_i \) are independent so that the \( n \) columns of \( U: n \times n \) are independent. Then \( U \) is of full rank and hence nonsingular. That is \( U^{-1} \) exists. Postmultiplying (A.3.1) by \( U^{-1} \) we obtain

\[
A = UDU^{-1}
\]

Define

\[
V = U^{-1}
\]

and premultiplying (A.3.3) by (A.3.4) we get

\[
VA = DV
\]

The rows \( v_i^T \) of \( V \) are therefore such that

\[
v_i^T A = \lambda_i v_i^T
\]
that is, the left eigenvectors and \( \lambda_i \) the eigenvalues.

Rewriting (A.3.3) using (A.3.4) we have
\[
A = UD \lambda V
\]
\[
= [u_1 \ u_2 \ \ldots \ u_n] \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & \lambda_n
\end{bmatrix}
\begin{bmatrix}
v_1 \\
v_2 \\
\vdots \\
v_n
\end{bmatrix}
\]
\[
= \sum_{i=1}^{n} \lambda_i u_i v_i^t
\]

Theorem A.4
Let \( X = [X_1 \ \ldots \ X_p]^t \) be a p-variate normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \). Then
\[
X^t X \sim \chi_p^2
\]

\textbf{Proof}

The density function of \( X \) is
\[
f(x) = \left( \frac{1}{\sqrt{2\pi}} \right)^p \exp \left( -\frac{1}{2} \sum_{i=1}^{p} x_i^2 \right)
\]
where
\[
x = [x_1 \ x_2 \ \ldots \ x_p]^t
\]

The mgf of \( X^t X \) is given by
\[
E[\exp(tX^tX)] = E[\exp(t \sum_{i=1}^{n} X_i^2)]
\]
\[
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \right)^p \exp \left[ -\frac{1}{2} \sum_{i=1}^{p} dz_i^2 \right] dz_1 \cdots dz_p
\]
\[
\begin{align*}
&= \frac{p}{\prod_{i=1}^{\infty}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(1-2t)z_i^2}{2}\right] dz_i : X_i \text{ are independent} \\
&= (1-2t)^{-p/2} \frac{p}{\prod_{i=1}^{\infty}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-2t)^{-1}}} \exp\left[-\frac{(1-2t)z_i^2}{2}\right] dz_i \\
&= (1-2t)^{-p/2} \prod_{i=1}^{p} 1: \text{Integral represents whole area of normal curve } N(0, (1-2t)^{-1}) \\
&= (1-2t)^{-p/2}
\end{align*}
\]

which is the mgf of \( \chi_p^2 \).

Therefore, by uniqueness of mgf,

\[
X'X \sim \chi_p^2
\]
### APPENDIX B

**DATA SETS**

**Data Set B.1**

Harvey (1989: 526)

Mink-muskrat furs sold by Hudson's Bay Company

#### Annual data 1848–1911

<table>
<thead>
<tr>
<th>Year</th>
<th>Mink</th>
<th>Muskrat</th>
<th>Mink</th>
<th>Muskrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1848</td>
<td>37133</td>
<td>224347</td>
<td>1881</td>
<td>36160</td>
</tr>
<tr>
<td>1849</td>
<td>34712</td>
<td>179075</td>
<td>1882</td>
<td>45600</td>
</tr>
<tr>
<td>1850</td>
<td>29610</td>
<td>115472</td>
<td>1883</td>
<td>47508</td>
</tr>
<tr>
<td>1851</td>
<td>21151</td>
<td>194682</td>
<td>1884</td>
<td>52290</td>
</tr>
<tr>
<td>1852</td>
<td>24859</td>
<td>292530</td>
<td>1885</td>
<td>110824</td>
</tr>
<tr>
<td>1853</td>
<td>25152</td>
<td>493952</td>
<td>1886</td>
<td>75833</td>
</tr>
<tr>
<td>1854</td>
<td>42352</td>
<td>512391</td>
<td>1887</td>
<td>64303</td>
</tr>
<tr>
<td>1855</td>
<td>50839</td>
<td>345626</td>
<td>1888</td>
<td>344878</td>
</tr>
<tr>
<td>1856</td>
<td>61851</td>
<td>258809</td>
<td>1889</td>
<td>46748</td>
</tr>
<tr>
<td>1857</td>
<td>58124</td>
<td>366244</td>
<td>1890</td>
<td>223614</td>
</tr>
<tr>
<td>1858</td>
<td>75233</td>
<td>173542</td>
<td>1891</td>
<td>47256</td>
</tr>
<tr>
<td>1859</td>
<td>63242</td>
<td>252246</td>
<td>1892</td>
<td>40748</td>
</tr>
<tr>
<td>1860</td>
<td>44730</td>
<td>177286</td>
<td>1893</td>
<td>35396</td>
</tr>
<tr>
<td>1861</td>
<td>31094</td>
<td>205020</td>
<td>1894</td>
<td>29479</td>
</tr>
<tr>
<td>1862</td>
<td>49652</td>
<td>335385</td>
<td>1895</td>
<td>24264</td>
</tr>
<tr>
<td>1863</td>
<td>43961</td>
<td>357069</td>
<td>1896</td>
<td>58171</td>
</tr>
<tr>
<td>1864</td>
<td>61727</td>
<td>505769</td>
<td>1897</td>
<td>50815</td>
</tr>
<tr>
<td>1865</td>
<td>60334</td>
<td>414577</td>
<td>1898</td>
<td>51285</td>
</tr>
<tr>
<td>1866</td>
<td>51424</td>
<td>232824</td>
<td>1899</td>
<td>70239</td>
</tr>
<tr>
<td>1867</td>
<td>58451</td>
<td>412164</td>
<td>1900</td>
<td>76865</td>
</tr>
<tr>
<td>1868</td>
<td>73575</td>
<td>619051</td>
<td>1901</td>
<td>70407</td>
</tr>
<tr>
<td>1869</td>
<td>74343</td>
<td>404173</td>
<td>1902</td>
<td>41839</td>
</tr>
<tr>
<td>1870</td>
<td>27708</td>
<td>232251</td>
<td>1903</td>
<td>48978</td>
</tr>
<tr>
<td>1871</td>
<td>31985</td>
<td>443899</td>
<td>1904</td>
<td>47813</td>
</tr>
<tr>
<td>1872</td>
<td>39266</td>
<td>704789</td>
<td>1905</td>
<td>57620</td>
</tr>
<tr>
<td>1873</td>
<td>31740</td>
<td>767996</td>
<td>1906</td>
<td>66549</td>
</tr>
<tr>
<td>1874</td>
<td>60429</td>
<td>671952</td>
<td>1907</td>
<td>54673</td>
</tr>
<tr>
<td>1875</td>
<td>72273</td>
<td>528002</td>
<td>1908</td>
<td>53596</td>
</tr>
<tr>
<td>1876</td>
<td>79214</td>
<td>583319</td>
<td>1909</td>
<td>66605</td>
</tr>
<tr>
<td>1877</td>
<td>59060</td>
<td>437121</td>
<td>1910</td>
<td>39769</td>
</tr>
<tr>
<td>1878</td>
<td>84244</td>
<td>486030</td>
<td>1911</td>
<td>21534</td>
</tr>
<tr>
<td>1879</td>
<td>62990</td>
<td>499727</td>
<td>1912</td>
<td>17857</td>
</tr>
<tr>
<td>1880</td>
<td>35072</td>
<td>48906</td>
<td>1913</td>
<td>21788*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>749154*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1911</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33008*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>963570*</td>
</tr>
</tbody>
</table>

* scores estimates from logarithms of same data in Data Set B.2
Data Set B.2
Reinsel (1993: 226)
Natural logarithms of annual sales of mink and muskrat furs sold by the Hudson's Bay Company for the years 1848-1911

<table>
<thead>
<tr>
<th>Year</th>
<th>Log Mink</th>
<th>Log Muskrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td>10.2962</td>
<td>12.6752</td>
</tr>
<tr>
<td>1851</td>
<td>9.9804</td>
<td>12.1701</td>
</tr>
<tr>
<td>1852</td>
<td>10.1210</td>
<td>12.5863</td>
</tr>
<tr>
<td>1853</td>
<td>10.1327</td>
<td>12.1192</td>
</tr>
<tr>
<td>1854</td>
<td>10.6543</td>
<td>13.1466</td>
</tr>
<tr>
<td>1855</td>
<td>10.8364</td>
<td>12.7531</td>
</tr>
<tr>
<td>1856</td>
<td>11.0281</td>
<td>12.4698</td>
</tr>
<tr>
<td>1857</td>
<td>11.0541</td>
<td>12.6191</td>
</tr>
<tr>
<td>1858</td>
<td>11.2415</td>
<td>12.6556</td>
</tr>
<tr>
<td>1859</td>
<td>11.0551</td>
<td>12.4461</td>
</tr>
<tr>
<td>1860</td>
<td>10.7084</td>
<td>12.0855</td>
</tr>
<tr>
<td>1861</td>
<td>10.3448</td>
<td>12.2357</td>
</tr>
<tr>
<td>1862</td>
<td>10.8038</td>
<td>12.7230</td>
</tr>
<tr>
<td>1863</td>
<td>10.6911</td>
<td>12.7857</td>
</tr>
<tr>
<td>1864</td>
<td>11.0305</td>
<td>13.1417</td>
</tr>
<tr>
<td>1865</td>
<td>11.0017</td>
<td>12.9441</td>
</tr>
<tr>
<td>1866</td>
<td>10.8475</td>
<td>12.6786</td>
</tr>
<tr>
<td>1867</td>
<td>10.9739</td>
<td>12.9292</td>
</tr>
<tr>
<td>1868</td>
<td>11.2061</td>
<td>13.3344</td>
</tr>
<tr>
<td>1869</td>
<td>11.2164</td>
<td>12.9806</td>
</tr>
<tr>
<td>1870</td>
<td>11.2629</td>
<td>12.5586</td>
</tr>
<tr>
<td>1871</td>
<td>10.3730</td>
<td>11.9026</td>
</tr>
<tr>
<td>1872</td>
<td>10.5781</td>
<td>13.4347</td>
</tr>
<tr>
<td>1873</td>
<td>10.7686</td>
<td>13.5514</td>
</tr>
<tr>
<td>1874</td>
<td>10.9092</td>
<td>14.3481</td>
</tr>
<tr>
<td>1875</td>
<td>11.1882</td>
<td>13.1689</td>
</tr>
<tr>
<td>1876</td>
<td>11.2700</td>
<td>12.2765</td>
</tr>
<tr>
<td>1877</td>
<td>11.3958</td>
<td>12.0840</td>
</tr>
<tr>
<td>1878</td>
<td>11.2702</td>
<td>11.8021</td>
</tr>
<tr>
<td>1879</td>
<td>11.5422</td>
<td>12.7318</td>
</tr>
<tr>
<td>1880</td>
<td>10.4652</td>
<td>13.0773</td>
</tr>
</tbody>
</table>

* scores calculated from Data Set B.1
BIBLIOGRAPHY


HARVEY, AC (1981a) The Econometric Analysis of Time Series; Philip Allan, Oxford
- (1981b) Time Series Models; Philip Allan, Oxford


