

Vector Autoregressions  
and Common Trends in  
Macro and Financial Economics



# STOCKHOLM SCHOOL OF ECONOMICS

The Economic Research Institute/EFI

Address: Sveavägen 65, Box 6501, S-113 83 Stockholm, tel 08-736 90 00

---

## Basic Orientation

The Economic Research Institute, at the Stockholm School of Economics (EFI) is devoted to the scientific study of problems in management science and economics. It provides research facilities for scholars belonging to these disciplines, training advanced students in scientific research. The studies to be carried out by the Institute are chosen on the basis of their expected scientific value and their relevance to the research programs outlined by the different sections of the Institute.

## Research Divisions:

- A Work Life Issues and Personnel
- B Accounting and Managerial Finance
- C Managerial Economics
- CFR Center for Risk Research
- D Marketing, Distribution and Industry Dynamics
- ES Economic Statistics
- F Public Administration
- FI Finance
- I Information Management
- IEG International Economics and Geography
- P Economic Psychology
- S Economics

## Independent Research Programs:

- PMO Man and Organization
- The Foundation for Distribution Research
- Research Program in Policy Sciences

---

Additional information about research in progress and published reports is described in our project catalogue. The catalogue can be ordered directly from The Economic Research Institute, Box 6501, S-113 83 Stockholm, Sweden.

# Vector Autoregressions and Common Trends in Macro and Financial Economics

Anders Warne



A Dissertation for the  
Doctors's Degree in Economics  
Stockholm School of Economics 1990

© EFI and the author

ISBN 91-7258-310-X

*Keywords:*

Vector autoregression  
Cointegration  
Impulse response function  
Variance decomposition  
Asymptotic theory  
Term structure

Gotab, Stockholm 1990

# Preface

This writing business. Pencils and what-not.  
Over-rated, if you ask me. Silly stuff. Nothing in it.

*Eyore*

In the early hours of a cold January morning in 1988 the angry signals from a phone caught my drowsy attention. Not entirely sure of my position in space I orbited my arm in the direction of the noise, brought the phone to the vicinity of my head and grunted a seven a.m. hello. From the other end of the line the enthusiastic voice of Peter Englund poured strange words and phrases into my confused head. I distinctly remember things like: "...prosearch reject ...coimplication ...serious analysis ..." and even greater complexities which I still don't know how to pronounce or how to spell for that matter. To end the phonecall as quickly as possible I said I was interested and hung up.<sup>1</sup>

Six months later I left Minneapolis for Stockholm and began to work on time series models with cointegration restrictions. At the time, my experience in this field was limited to having been present at workshops in Helsingør, Denmark, and Minneapolis, Minnesota, where people like Clive Granger, David Hendry and Christopher Sims talked about some of its basic ideas, implications and shortcomings. Nevertheless, had it not been for my collaboration with Peter Englund, Erik Mellander and Anders Vredin this thesis would undoubtedly have another title and deal with something completely different.

In the course of time I have also benefitted greatly from advice, discussions and correspondence with Michael Bergman, Nils Gottfries, Niels Haldrup, Tor Jacobson, Søren Johansen, Karl Jungentfelt, Katarina Juselius, Sune Karlsson, Helmut Lütkepohl, Lars-Erik Öller, Aris Protopapadakis, Avi Simhony and Karl Wärneryd. I am especially grateful to my thesis adviser, Staffan Viotti (Y80), my computer consultant, Ossian Ekdahl, my high school teacher Hans Ekroth for getting me interested in economics, my versatile tutor, the late Fred Wilson, and my brother-in-arms and applied game theory expert, Dave Christian. I am also grateful to the staff, researchers and fellow students at the Department of Economics, University of Minnesota, Minneapolis, the Department of Financial Economics, Stockholm School of Economics, and FIEF, Stockholm. A very special thanks to the weekend crew: Catharina Lagerstam, Kerstin Lindskog and Tomas Vaclavinek.

---

<sup>1</sup>In a popular Soviet joke, Comrade Leader and Comrade General are reviewing the forces of the May Day parade in Red Square. At the end of the parade, behind the last missile, walk two men. "And who are those two?" asks Comrade Leader. "Ah," replies Comrade General, "they are economists." "Why are they in the parade?" asks Comrade Leader. "Well," Comrade General replies jubilantly, "nothing equals the destructive power of *two* economists."

For inspiration I wish to thank M. Camilo, Brita and Oskar, Sergei P., Scott H., Reynolds and Camels, Charlie P., Droopy, M. Roberts, Grandma's, B. Berg and M. Stern, G. Vladimov, Zdena T., Chick C., B. Fields, B. Adder, Miles D., M. Groening, Big G., Double R., B. Breathed, C. Hunter, H. and A. Selmer, J. Patitucci, Opus, Wolfgang M., D. Weckl, Rosie O'Grady's, R.L. Jones, Maurice R., A. Gaudi, Te Mutunga – Ranei Te Take, G. Larson, Antonin D., Kirby P., D. Malouf, C. Kent, Sebastian B., K. Hulme, F. Gambale for legato licks, J.M. Coetzee, L. George and L. Feat, Snell A., Milan K., the Pooh, John C., Basil F., E. Marienthal, S. Dan, A. Ghosh, Reb's, Jaroslav H., all the Goede's in the world, Michael and Randy B., Nols P., Mutts G., Larry L., etc. Finally, I owe it all to my family: from Greenpeace supporters to missile design engineers, from up and coming journalists to elementary philosophers, from Birdlovers to Mother of punk, from rockin' horse cousins to furry woodpeckers, Yanks, Brits, Dasswedanyans, Scots, ex-Aussies and the lot – in one way or another you are all to blame; to Scotty for Warp II, an occasional Warp III, and a rare imitation of Hal (whenever the Klingons invade my programs); to T<sub>E</sub>X for typesetting; and to Alyoshie for getting things done.

Financial support from Jan Wallanders Research Foundation, the Royal Academy of Sciences (Karl Langenskiölds Foundation), the Stockholm School of Economics (Louis Fraenckels Foundation), Humanistisk-Samhällsvetenskapliga Forskningsrådet and Bank-forskningsinstitutet is gratefully acknowledged.

Stockholm, October 1990,

Anders Warne

# Contents

<b>1</b>	<b>Vector Autoregressions and Cointegration</b>	<b>1</b>
1.1	Some Useful Definitions . . . . .	1
1.2	A Summary of the Thesis . . . . .	5
<b>2</b>	<b>A Common Trends Model: Representation, Estimation and Asymptotic Properties</b>	<b>7</b>
2.1	Introduction . . . . .	7
2.2	Estimation of a Common Trends Model . . . . .	8
2.2.1	A Mathematical Structure for Cointegrated Time Series . . . . .	9
2.2.2	Estimation of the Common Trends Parameters . . . . .	12
2.2.3	An Example . . . . .	14
2.3	Inversion and Identification . . . . .	15
2.3.1	The Moving Average Parameters . . . . .	15
2.3.2	Identification of Permanent and Transitory Innovations . . . . .	17
2.4	Consistency and Asymptotic Normality . . . . .	20
2.4.1	Asymptotic Properties of Impulse Response Functions . . . . .	20
2.4.2	Derivation of the Forecast Error Variance Expressions . . . . .	29
2.4.3	Asymptotic Properties of Forecast Error Variance Decompositions . . . . .	31
2.5	Concluding Comments . . . . .	36
<b>3</b>	<b>Common Trends and Macroeconomic Fluctuations</b>	<b>38</b>
3.1	Introduction . . . . .	38
3.2	Analysis of Macroeconomic Fluctuations . . . . .	38
3.3	Cointegration Vectors and Common Trends in a Small Open Economy Growth Model . . . . .	41
3.4	Common Trends and Macroeconomic Fluctuations in Finland and Sweden . . . . .	45
3.4.1	The Data . . . . .	45
3.4.2	Illustrations . . . . .	46
3.5	Concluding Remarks . . . . .	61
<b>4</b>	<b>Stationarity, Cointegration and a Rational Expectations Model of the Term Structure</b>	<b>62</b>
4.1	Introduction . . . . .	62
4.2	An Expectations Model of the Term Structure . . . . .	63
4.2.1	The Local Expectations Hypothesis . . . . .	64
4.2.2	Discussion of the Model . . . . .	65
4.3	Connecting Economic Theory with Data . . . . .	66

4.3.1	The Data Generating Model . . . . .	66
4.3.2	Term Structure Restrictions . . . . .	68
4.3.3	Implications for Statistical Analysis . . . . .	71
4.3.4	A Joint Test of Cointegration and the Term Structure Model . . . . .	73
4.4	Empirical Results . . . . .	77
4.4.1	The Data . . . . .	77
4.4.2	Tests of the Term Structure Model in Swedish Bond Markets . . . . .	77
4.4.3	Inference from Bootstrap Distributions . . . . .	85
4.5	Concluding Remarks . . . . .	90

# Chapter 1

## Vector Autoregressions and Cointegration

### 1.1 Some Useful Definitions

Vector autoregressive models have primarily been applied to economic data for three purposes:

- (i) to forecast macroeconomic indicators (see, e.g., Doan, Litterman and Sims [21], and Litterman [55]),
- (ii) to study the sources and characteristics of economic fluctuations (see, e.g., Cooley and LeRoy [17], Keating [50], and Sims [90,91]), and
- (iii) to test economic theories which specify relationships between the present and the expected future realisations of a set of variables (see, e.g., Baillie [4], and Sargent [81]).

In the following three chapters I shall consider the latter two areas. Specifically, I will analyse vector autoregressions with cointegration constraints (cf. Engle and Granger [26], and Granger and Weiss [34]).

The basic idea of cointegration is very simple: although a vector time series may have the property that each individual series is nonstationary certain linear combinations of such processes can be stationary. For example, aggregate income and consumption are generally not well modelled as wide sense stationary. However, the permanent income hypothesis may suggest that these two series do *not* drift too far apart (see, e.g., Campbell [13] and Stock and Watson [96]).

From a practical point of view there are several reasons why cointegration is of interest. In relation to (ii) above, cointegrating or steady state relationships can be used to identify stochastic trends in the data. Innovations to such trends may or may not be an important factor for describing short run fluctuations in macroeconomic variables such as output, the price level, investments, etc. In other words, cointegration provides a tool for modelling nonstationarity by imposing steady state or long run equilibrium relationships on the variables. Furthermore, the standard distribution theory for hypothesis testing in a time series framework relies on the assumption that data are stationary. If this assumption is not satisfied then inference will generally be based on the wrong limiting distributions (see,

e.g., Sims, Stock and Watson [92]). In this case, possible cointegrating relationships can often be derived from the economic hypothesis. Once we have conditioned the econometric or time series model on these restrictions we may be back in a setting where standard asymptotic theory is applicable. For example, the market efficiency hypothesis of exchange rates states that the expected one period ahead spot exchange rate is equal to the current one period forward rate. If exchange rates are nonstationary in levels and stationary in changes the market efficiency hypothesis suggests that the spread between the forward and spot exchange rates is stationary.

To be more precise about the notion of cointegration, let  $\{x_t\}_{t=1}^{\infty}$  be an  $n$  dimensional real valued stochastic process generated according to the vector autoregression:

$$x_t = \rho + \sum_{k=1}^p A_k x_{t-k} + \varepsilon_t, \quad (1.1)$$

where  $\{x_0, \dots, x_{1-p}\}$  are  $np$  (deterministic) initial conditions,  $\varepsilon_t$  is purely nondeterministic and serially uncorrelated with mean zero and positive definite covariance matrix  $\Sigma$ , and with  $\varepsilon_s = 0$  for all  $s \leq 0$ . Also, let  $E[x_t] = \mu_t$  and  $\tilde{x}_t := x_t - \mu_t$  for all  $t \geq 1$ , where  $:=$  denotes a definition.

**Definition 1.1** A vector time series  $\{x_t\}_{t=1}^{\infty}$  is said to be jointly wide sense stationary if (i)  $\mu_t = \mu$ , and (ii)  $E[(x_t - \mu)(x_{t-k} - \mu)'] = \Upsilon_k$  for all  $t$  and  $k$ .

To avoid ambiguities below consider also the following:

**Definition 1.2** A vector time series  $\{x_t\}_{t=1}^{\infty}$  is said to be jointly covariance or second moment stationary if  $\{\tilde{x}_t\}_{t=1}^{\infty}$  is jointly wide sense stationary.

It should be noted that if  $x_t$  is generated according to

$$x_t = \delta t + \varepsilon_t, \quad (1.2)$$

then  $\{x_t\}_{t=1}^{\infty}$  is jointly covariance but *not* jointly wide sense stationary. Furthermore, subtracting  $x_{t-1}$  from both sides of equation (1.2) we find that

$$x_t - x_{t-1} = \delta + \varepsilon_t - \varepsilon_{t-1}. \quad (1.3)$$

Accordingly,  $\{(x_t - x_{t-1})\}_{t=1}^{\infty}$  is jointly wide sense stationary. In fact, this exemplifies why the following definition of integration is relevant:

**Definition 1.3** If  $\{x_t\}_{t=1}^{\infty}$  is not second moment stationary but  $\{(x_t - x_{t-1})\}_{t=1}^{\infty}$  is jointly wide sense stationary, then  $\{x_t\}_{t=1}^{\infty}$  is said to be integrated of order 1.

Thus, if the stochastic process  $\{x_t\}_{t=1}^{\infty}$  is jointly covariance stationary it cannot be integrated of order one (and vice versa) although  $\{x_t\}_{t=1}^{\infty}$  in (1.2) is *not* wide sense stationary while  $\{(x_t - x_{t-1})\}_{t=1}^{\infty}$  is. This motivates the condition that  $\{x_t\}_{t=1}^{\infty}$  should not be second moment stationary in Definition 1.3. It can also be noted that if the sequence  $\{x_t\}_{t=1}^{\infty}$  is jointly wide sense stationary I will also refer to it as being integrated of order zero.

**Definition 1.4** If  $\{x_t\}_{t=1}^{\infty}$  is integrated of order 1 and for some  $\alpha \in \mathcal{R}^{n \times r}$  with  $\text{rank}[\alpha] = r < n$  the  $r$  dimensional stochastic process  $\{z_t\}_{t=1}^{\infty}$ ,  $z_t := \alpha' x_t$ , is jointly wide sense stationary, then  $\{x_t\}_{t=1}^{\infty}$  is said to be cointegrated of order (1,1) with  $r$  cointegration vectors.

The terminology here follows that in Engle and Granger [26] although Definition 1.4 differs slightly from their definition of cointegrated processes of order (1,1). In fact, the above definition is more general than Engle and Granger's since it allows some elements of  $x_t$  to be wide sense stationary while Engle and Granger require all elements to be integrated of order one. From a practical perspective Definition 1.4 is preferable since we may be interested in studying a system where we wish to model some variables (e.g., unemployment) as wide sense stationary and others as integrated of order one (e.g., real output). Note that the first 1 in the order term refers to  $\{x_t\}_{t=1}^{\infty}$  being integrated of order 1, while the second 1 is obtained from  $\{z_t\}_{t=1}^{\infty}$  being integrated of order  $1 - 1 = 0$ , i.e., the cointegration vectors decrease the order of integration by one.

For the stochastic process described in (1.2) there exists an  $n \times (n - 1)$  matrix  $\alpha$  with rank  $(n - 1)$  such that  $\alpha'\delta = 0$ , i.e.,  $\{\alpha'x_t\}_{t=1}^{\infty}$  is jointly wide sense stationary. However, since  $\{x_t\}_{t=1}^{\infty}$  is *not* integrated of order one it follows that this stochastic process is not cointegrated of order (1,1) with  $(n - 1)$  cointegration vectors. Furthermore, this model cannot be written in the form of equation (1.1).

To illustrate what is implied by these definitions, suppose the stochastic process generated by (1.1) is two dimensional with  $x_0 = 0$  and

$$\begin{aligned} x_{1,t} &= x_{1,t-1} + \varepsilon_{1,t}, \\ x_{2,t} &= x_{1,t-1} + \varepsilon_{2,t}. \end{aligned} \tag{1.4}$$

Solving the first equation we find that

$$x_{1,t} = \sum_{s=1}^t \varepsilon_{1,s}, \tag{1.5}$$

and substituting this into the second equation of (1.4) for  $x_{1,t-1}$  we obtain

$$x_{2,t} = \sum_{s=1}^{t-1} \varepsilon_{1,s} + \varepsilon_{2,t}. \tag{1.6}$$

The two dimensional vector time series  $\{x_t\}_{t=1}^{\infty}$  is integrated of order one. To show this, it can first be seen that  $E[x_t] = 0$ , while the contemporaneous covariance matrix of  $x_t$  is given by

$$E[x_t x_t'] = \begin{bmatrix} t\sigma_{11} & (t-1)\sigma_{11} + \sigma_{12} \\ (t-1)\sigma_{11} + \sigma_{12} & (t-1)\sigma_{11} + \sigma_{22} \end{bmatrix},$$

where  $\sigma_{ij}$  is the  $(i, j)$ :th element of  $\Sigma$ .<sup>1</sup> Hence, second moments of  $x_t$  are functions of  $t$  and  $\{x_t\}_{t=1}^{\infty}$  is thus not covariance stationary in levels. The first difference or change of  $x_t$  is

$$\begin{aligned} x_{1,t} - x_{1,t-1} &= \varepsilon_{1,t}, \\ x_{2,t} - x_{2,t-1} &= \varepsilon_{2,t} + \varepsilon_{1,t-1} - \varepsilon_{2,t-1}. \end{aligned} \tag{1.7}$$

The contemporaneous covariance matrix of  $(x_t - x_{t-1})$  is equal to

$$E[(x_t - x_{t-1})(x_t - x_{t-1})'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} + 2(\sigma_{22} - \sigma_{12}) \end{bmatrix}.$$

---

<sup>1</sup>It can be noted that  $\lim_{t \rightarrow \infty} E[t^{-1}x_t x_t']$  converges to a singular matrix with  $\sigma_{11}$  in each of its elements. From equations (1.5)–(1.6) it can be seen that  $x_t$  has one common stochastic trend. It is precisely the variance of this trend that dominates the uncertainty of  $x_t$ .

Neither this matrix nor the autocovariance matrices  $E[(x_t - x_{t-1})(x_s - x_{s-1})']$  are functions of  $t$  or  $s$ . Accordingly,  $\{x_t\}_{t=1}^{\infty}$  is jointly wide sense stationary in first differences.

Furthermore,  $x_t$  has one cointegration vector which can be described by  $\alpha' = [1 \quad -1]$  so that

$$z_{1,t} := x_{1,t} - x_{2,t} = \varepsilon_{1,t} - \varepsilon_{2,t}, \quad (1.8)$$

is wide sense stationary since its mean is constant,  $E[z_{1,t}^2] = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$ , whereas  $E[z_{1,t}z_{1,s}] = 0$  for all  $s \neq t$ , i.e.,  $\{z_{1,t}\}_{t=1}^{\infty}$  is a white noise process.

Looking at these equations it can be seen that if we suppose the parameters are unknown, then only the system in (1.4) is easily estimated. The other two systems, i.e., (1.5)–(1.6) and (1.7), involve serially correlated residuals. I shall refer to the former system as a stochastic trends (or reduced form common trends) representation, and to the latter as a vector moving average representation. Note, however, that if (1.4) is estimated the fact that  $z_{1,t}$  in (1.8) is wide sense stationary will not be taken into account in finite samples (see, e.g., Engle and Yoo [28]). If  $\alpha$  is known, we may instead consider the following two representations

$$\begin{aligned} x_{1,t} - x_{1,t-1} &= \varepsilon_{1,t}, \\ x_{2,t} - x_{2,t-1} &= z_{1,t-1} + \varepsilon_{2,t}, \end{aligned} \quad (1.9)$$

and

$$\begin{aligned} x_{1,t} - x_{1,t-1} &= \varepsilon_{1,t}, \\ z_{1,t} &= \varepsilon_{1,t} - \varepsilon_{2,t}. \end{aligned} \quad (1.10)$$

The first system is called a vector error correction representation since the changes in  $x$  at time  $t$  reacts to a steady state error ( $x_1 - x_2$ ) at time  $t - 1$ . The second system is called a restricted vector autoregression as it is expressed in vector autoregressive form but is conditioned on the given cointegration relationship. Both these representations of (1.4) take the assumed form of nonstationarity into account and are easy to estimate since the unobserved variables are serially uncorrelated.

It should be emphasized that if  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with  $r$  cointegration vectors, then all elements of this stochastic process need *not* be nonstationary. To illustrate this, consider the following unrestricted vector autoregression:

$$\begin{aligned} x_{1,t} &= x_{1,t-1} + \varepsilon_{1,t}, \\ x_{2,t} &= x_{1,t-1} + \varepsilon_{2,t}, \\ x_{3,t} &= x_{1,t-1} - x_{2,t-1} + \varepsilon_{3,t}. \end{aligned} \quad (1.11)$$

The first two equations correspond to the earlier system (cf. equation (1.4)). The third equation specifies that  $x_{3,t}$  is equal to the difference between  $x_1$  and  $x_2$  at  $t - 1$  plus a white noise residual. In terms of a restricted vector autoregression the system in (1.11) can be written as

$$\begin{aligned} x_{1,t} - x_{1,t-1} &= \varepsilon_{1,t}, \\ z_{1,t} &= \varepsilon_{1,t} - \varepsilon_{2,t}, \\ z_{2,t} &= z_{1,t-1} + \varepsilon_{3,t}. \end{aligned} \quad (1.12)$$

Consequently, there are now two cointegration vectors which can be described by the matrix

$$\alpha' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For this case there are two series,  $\{x_{1,t}\}_{t=1}^{\infty}$  and  $\{x_{2,t}\}_{t=1}^{\infty}$ , which are integrated of order one, and one series,  $\{x_{3,t}\}_{t=1}^{\infty}$ , which is integrated of order zero. As a system, however, we find that  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with two cointegration vectors and is therefore nonstationary in levels.

In the theoretical and empirical analyses of financial and macroeconomic time series below I shall concentrate on the restricted vector autoregression. There are at least two reasons for this. First, it is autoregressive and standard theory on  $s$  steps ahead forecasts (“rational expectations”), estimators, asymptotic properties, multivariate tests, etc., can thus be applied. Second, the cointegration and common trends literature is mainly concerned with the vector error correction representation.<sup>2</sup> Hence, its pros and cons are already fairly well known.<sup>3</sup>

## 1.2 A Summary of the Thesis

In Chapter 2 necessary and sufficient conditions for the existence of a restricted vector autoregression are given. Furthermore, it is shown how we can estimate a common trends model from a restricted vector autoregression when the  $n$  dimensional vector time series  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with  $r$  cointegration vectors. Identification of  $k = n - r$  permanent and  $r$  transitory innovations for estimating impulse response functions and forecast error variance decompositions is analysed in detail. Analytical expressions of the asymptotic distributions for estimates of these functions are derived for the case when  $\alpha$  and an upper bound for  $p$ , the lag order, are known. Here, the property that the sum of a forecast error variance decomposition is one is explicitly taken into account. This has, to my knowledge, not been dealt with previously in the literature. The chapter is based on a paper by myself with the title “Estimating and Analysing the Dynamic Properties of a Common Trends Model”.

Chapter 3 is concerned with an illustration of how the theoretical results in Chapter 2 can be used to analyse macroeconomic fluctuations in small open economies. The data I use is annual from Sweden (1871–1986) and Finland (1866–1985). It should be emphasized that it is not my intention here to argue that a common trends model is well suited for analysing these particular samples. Instead the motivation for the exercises in this chapter is to make the rather technical material in Chapter 2 more accessible to economists who are mainly interested in empirical analysis of macroeconomic data. For each economy I identify three common stochastic trends in a vector of six variables containing terms of trade, the price level, the money stock, real output, real investments, and real consumption. The cointegration vectors are derived from a steady state solution to a small open economy real business cycle model and the trends are interpreted as a

---

<sup>2</sup>See, e.g. Johansen [44, Theorem 4.1] for conditions under which  $\{x_t\}_{t=1}^{\infty}$  in (1.1) is cointegrated of order (1,1) with  $r$  cointegration vectors.

<sup>3</sup>It may also be added that a vector autoregression, in my opinion, has a more beautiful mathematical composition than a vector error correction model.

foreign, a real domestic, and a nominal domestic trend. Shocks to these trends are then examined via impulse response functions and forecast error variance decompositions. The chapter is based on a joint paper with Erik Mellander, the Industrial Institute for Economic and Social Research (IUI), and Anders Vredin, Trade Union Institute for Economic Research (FIEF), with the title “Stochastic Trends and Economic Fluctuations in Small Open Economies: The Cases of Finland and Sweden”.

Finally, in Chapter 4 I analyse the connections between a rational expectations model of the term structure of interest rates and a vector autoregressive data generating model. For nonexplosive time series it is shown that the long term yield series Granger-causes the short term yield when data are consistent with the economic model. Furthermore, the term structure model generally implies that jointly wide sense stationary interest rates are *nearly* cointegrated. A simple likelihood ratio test is proposed to examine cointegration and economic theory jointly. From bootstrap and  $\chi^2$  based inference on Swedish data on 1-month Treasury bills and 5-year Treasury bonds (1983:11–1989:12) it is found that possible nonstationarity probably does not need to be accounted for when testing the term structure model, and that the  $\chi^2$  distribution performs well in relation to the bootstrap distributions as estimates of the small sample uncertainty. The chapter is based on joint work (same title) with Ossian Ekdahl at the Department of Financial Economics, Stockholm School of Economics. Naturally, my co-authors are *not* responsible for any errors, obscurities or views expressed here. That burden is carried by me alone.

All numerical calculations in Chapters 3 and 4 have been performed by RATS, version 3.01. Interactive programs for all computations except the Johansen likelihood ratio tests have been written by myself and are available on request. Programs for performing the Johansen tests have been written by Søren Johansen, Katarina Juselius and Henrik Hansen at the University of Copenhagen.

## Chapter 2

# A Common Trends Model: Representation, Estimation and Asymptotic Properties

### 2.1 Introduction

In many models on macroeconomic fluctuations the dichotomy between growth and cycles has played an important role. Traditionally, growth has often been treated as deterministic in studies on business cycles (cf. King, Plosser and Rebelo [51]). In contrast, stochastic growth models, such as the real business cycle models considered by, i.a., King, Plosser and Rebelo [52], and King, Plosser, Stock and Watson [53], allow shocks to growth to influence the short run fluctuations.

A common feature of stochastic growth models is that the number of growth disturbances is rather low relative to the number of variables. The prevailing view in the theoretical literature seems to be that macroeconomic fluctuations arise from shocks to fundamental variables such as economic policy, preferences, and technology. These shocks are then propagated through the economy and result in systematic patterns of persistence and comovements among macroeconomic aggregates. Consequently, it should be of interest to analyse a simple time series model which makes it possible to examine connections between growth related shocks and transient fluctuations. Such a model will then by necessity incorporate stochastic rather than deterministic trends. Furthermore, to consider the notion of a few important growth disturbances, there will in general be fewer stochastic trends than time series.

In papers by King, Plosser, Stock, and Watson [53] and Stock and Watson [95], the connection between cointegration and common stochastic trends was first examined in some detail. The basic idea is that there is a *reduced* number of linear stochastic trends feeding the system. This implies that there exists certain linear combinations of the (log) levels series which ensure that the trends average out, i.e., the residuals from the linear combinations are wide sense stationary stochastic processes. King, Plosser, Stock, and Watson investigate a common trends model on five U.S. macroeconomic time series (output, consumption, investments, the price level, and the money stock) and model growth by two stochastic trends, a nominal and a real trend. With five time series and two trends, common sense (or algebra) suggests that we can construct three independent vec-

tors which eliminate the trends, i.e., there are three cointegrating vectors which describe a steady state in such a system. A shortcoming of their paper is that the description of the estimation and computation strategy they make use of is somewhat limited. For example, an inversion algorithm needed to obtain estimates of, e.g., impulse response functions and forecast error variance decompositions is only mentioned. Furthermore, asymptotic properties of these functions are not considered.

A purpose of this chapter is to mathematically establish how one may estimate the parameters in a common stochastic trends model when the time series of interest are cointegrated of order (1,1).<sup>1</sup> Furthermore, I shall show how one may perform dynamic analysis within this framework when the innovations to the system are either permanent or transitory, i.e., when the responses in at least one variable to an innovation are *or* are not persistent. In particular, the calculation of impulse response functions and forecast error variance decompositions will be looked into in some detail. Finally, I shall derive asymptotic distributions of estimates of these functions in the present setting. Here, the theory is based on Baillie [5,6], Lütkepohl [58,59,60], Lütkepohl and Poskitt [61], Lütkepohl and Reimers [62], and Schmidt [82,83], although the particular innovations I examine complicate the analysis somewhat. Furthermore, the fact that the sum of an  $s$  steps ahead forecast error variance decomposition is equal to one will be taken into account. To my knowledge, this has not been dealt with before.

The chapter is organized as follows. In section 2.2, I discuss some representations which are equivalent for cointegrated time series. There it is shown that a restricted vector autoregressive representation for cointegrated time series exists under familiar circumstances. Since this representation is invertible, it is well suited for calculating all other parameters of interest (see also Warne [98]). Section 2.3 is concerned with the moving average parameters and identification of permanent and transitory innovations. In section 2.4, I analyse the asymptotic properties of impulse response functions and forecast error variance decompositions under the assumptions that the cointegrating vectors are known and the lag order has a known upper bound. Section 2.5 summarizes the main results in the chapter. An illustration of the theory is discussed in Chapter 3.

## 2.2 Estimation of a Common Trends Model

Linear time series models which are conditioned on cointegrating restrictions are generally specified in terms of variables which can be observed and a purely nondeterministic and serially uncorrelated error. Accordingly, they can be estimated with standard tools. In contrast, a common trends model consists of a vector of trends and a vector of stationary variables, where neither component can be observed as an individual factor. Without loss of generality, let  $\{x_t\}_{t=1}^{\infty}$  be a vector time series such that

$$x_t = x_t^p + x_t^s.$$

Here,  $x_t^p$  represents a vector of trends of  $x_t$ , while  $x_t^s$  is a stationary residual. To estimate such a time series model it is therefore necessary to model the structure of either component and thereby identify both.

---

<sup>1</sup>See Blanchard and Quah [10], and Shapiro and Watson [86] for approaches which are related to the one I shall examine here.

In King, Plosser, Stock and Watson [53] and Stock and Watson [95] it is shown that there is a simple duality between the concepts of cointegration and common trends. In particular, the cointegrating restrictions determine the number of independent trends and how a vector of observed variables is related to all the independent trends. That is, if  $\alpha$  is a vector of cointegrating restrictions, then  $\alpha'x_t^p = 0$  for  $\alpha'x_t = \alpha'x_t^s$  to be stationary. These restrictions, however, generally *neither* specify *nor* suggest whether a certain trend is related to, e.g., technology shocks or economic policy. To be able to make such interpretations it is necessary to consider further identifying assumptions. In this section I shall devote the first part to the mathematical structure of cointegrated time series and the second part to estimation and identification of the common trends parameters. As an illustration, the third part contains an example of a three dimensional system for output, the price level, and the money stock.

## 2.2.1 A Mathematical Structure for Cointegrated Time Series

From Granger's Representation Theorem (cf. Engle and Granger [26], Hylleberg and Mizon [42], Johansen [44,45,46], and Warne [98]) we know that the following *vector moving average* model may be rewritten in various forms (representations) for an  $n$  dimensional real valued vector time series  $\{x_t\}_{t=1}^{\infty}$  which is cointegrated of order (1,1) with cointegrating rank  $r < n$ :

$$\Delta x_t = \delta + C(L)\varepsilon_t. \quad (2.1)$$

Here,  $L$  is the lag operator, i.e.,  $L^k x_t = x_{t-k}$  for any integer  $k$ ,  $\Delta := 1 - L$  is the first difference operator, the  $n \times n$  matrix polynomial  $C(\lambda) := I_n + \sum_{j=1}^{\infty} C_j \lambda^j$  is finite for all  $\lambda$  on or inside the unit circle,  $I_n$  is the  $n \times n$  identity matrix,  $\delta = C(1)\rho$  is the mean for  $\Delta x_t$ ,  $\text{rank}[C(1)] = n - r$ , and the  $n$  dimensional random variable  $\varepsilon_t$  is purely nondeterministic and serially uncorrelated with mean zero and positive definite covariance matrix  $\Sigma$ . Hence, cointegration implies, among other things, that  $C(1)$  has reduced rank so that  $C(\lambda)$  cannot be inverted with standard techniques.

From the above references, however, we find that the vector moving average representation in equation (2.1) is a solution to the following *unrestricted vector autoregressive* system:

$$A(L)x_t = \rho + \varepsilon_t, \quad (2.2)$$

where  $A(\lambda) = I_n - \sum_{k=1}^{\infty} A_k \lambda^k$ ,  $\text{rank}[A(1)] = r$ ,  $A(1) = \gamma\alpha'$ ,  $A(\lambda)C(\lambda) = C(\lambda)A(\lambda) = (1 - \lambda)I_n$ , the matrices  $\alpha$  and  $\gamma$  are of dimension  $n \times r$ , and the columns of  $\alpha$  are called the cointegrating vectors. From these properties it can be seen that equation (2.1) is a solution to (2.2).

Without loss of generality, let us assume that the matrix polynomial  $A(\lambda)$  is of finite order  $p$ , i.e.,  $A_{p+s} = 0$  for all  $s \geq 1$ . We can easily rewrite equation (2.2) as a *vector error correction* representation of the form

$$A^*(L)\Delta x_t = \rho - \gamma z_{t-1} + \varepsilon_t, \quad (2.3)$$

where the  $r$  dimensional vector time series  $\{z_t\}_{t=1}^{\infty}$  is integrated of order zero,  $z_t := \alpha'x_t$ ,  $A^*(\lambda) = I_n - \sum_{k=1}^{p-1} A_k^* \lambda^k$ , and  $A_k^* = -\sum_{l=k+1}^p A_l$ . In relation to equation (2.2), the vector error correction representation is based on the cointegrating matrix, i.e., all variables in the model are written in terms of stationary transformations.

In, e.g., Hylleberg and Mizon [42], King, Plosser, Stock, and Watson [53], Stock and Watson [95], and Warne [98] it is shown that there exists a *common trends* model of the form

$$x_t = x_0 + A\tau_t + \tilde{C}(L)\varepsilon_t, \quad (2.4)$$

where  $\tau_t$  is a  $k \times 1$  vector of random walks with drift  $\mu$  and innovation  $\varphi_t$ , i.e.,

$$\tau_t = \mu + \tau_{t-1} + \varphi_t. \quad (2.5)$$

It can be shown that  $k = n - r$ . Below I shall analyse dynamic effects from shocks to the  $k$  common trends. For that reason it is convenient to assume that the innovations,  $\varphi_t$ , are mutually independent so that ( $E[\varphi_t] = 0$  and)  $E[\varphi_t\varphi_t'] = \Phi = \text{diag}[\Phi_{11} \cdots \Phi_{kk}]$ .

The matrix  $A$  in equation (2.4) is  $n \times k$  with rank  $k$ , and the matrix polynomial  $\tilde{C}(\lambda) = \sum_{i=0}^{\infty} \tilde{C}_i \lambda^i$ . It can be shown that  $\tilde{C}_i = -\sum_{j=i+1}^{\infty} C_j$  (cf. Stock [94]). The condition that  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with cointegrating rank  $r$  implies that  $\tilde{C}(\lambda)$  is finite for all  $|\lambda| \leq 1$  and  $\text{rank}[\tilde{C}(1)] \geq r$ . To see this, note that the sequence  $\{z_t\}_{t=1}^{\infty}$ , where

$$z_t := \alpha'x_t = \alpha'x_0 + \alpha'\tilde{C}(L)\varepsilon_t, \quad (2.6)$$

is integrated of order zero. Below it will be shown that  $\alpha'\tilde{C}(L)\varepsilon_t$  represents an invertible vector moving average representation for  $z_t$  (cf. Theorem 2.1). For now it suffices to note that this can only be the case if  $\tilde{C}(\lambda)$  is finite and  $\tilde{C}(1)$  has rank at least equal to  $r$ . Furthermore, since the cointegrating (or steady state) relationships imply that  $\alpha'x_t^p = 0$  and  $x_t^p = A\tau_t$ , we have that  $\alpha'A = 0$ . Note also that  $x_0$  cannot be nonstationary. To simplify matters, it is assumed that  $x_0$  is constant.

The exact relationship between the the vector moving average and common trends model in equations (2.1) and (2.4) may now be established. As noted by, e.g., Stock [94] the matrix polynomial  $C(\lambda)$  can be written as  $C(\lambda) = C(1) + (1 - \lambda)\tilde{C}(\lambda)$ . Inserting this into equation (2.1), moving  $x_{t-1}$  to the right hand side, recursively substituting for  $x_{t-1}, \dots, x_1$ , and letting  $\varepsilon_s = 0$  for all  $s \leq 0$ , we obtain the following *stochastic trends* or reduced form common trends representation:

$$x_t = x_0 + C(1)\xi_t + \tilde{C}(L)\varepsilon_t, \quad (2.7)$$

where  $\xi_t = \rho + \xi_{t-1} + \varepsilon_t$ . The mathematical intuition for the reduced number of random walks in (2.4) stems from the fact that the matrix  $C(1)$  has rank  $n - r$ . Hence, only  $n - r$  elements of the  $n$  dimensional vector  $C(1)\varepsilon_t$  result in independent permanent effects on  $x_t$ . In fact, comparing (2.4), (2.6) and (2.7) we find that  $\alpha'A = \alpha'C(1) = 0$ . Furthermore, combining equations (2.4), (2.5), and (2.7) we get

$$A\varphi_t = C(1)\varepsilon_t, \quad A\Phi A' = C(1)\Sigma C(1)', \quad \text{and} \quad A\mu = C(1)\rho. \quad (2.8)$$

When one is concerned with estimating the common trends model, it is clear that we need to have information about the parameters of  $C(1)$ ,  $\Sigma$ , and  $\rho$ .<sup>2</sup>

---

<sup>2</sup>King, Plosser, Stock, and Watson [53] make use of the vector error correction representation in their study. Based on the results in Theorem 2.1 it can be shown that  $C(\lambda) = M^{-1}D(\lambda)[A^*(\lambda)M^{-1}D(\lambda) + N\lambda]^{-1}$ . An inversion routine which is built on this relationship seems somewhat more complex than the one derived in section 2.3 although they are algebraically equivalent; see also Lütkepohl and Reimers [62].

Campbell and Shiller [15] show that it is straightforward to rewrite the vector error correction representation as a restricted vector autoregressive system when  $n = 2$  and  $r = 1$ . Whereas Campbell [13] only discusses the matter, Warne [98, Theorem 2] shows that this result can be generalized to any choice of  $n$  and  $r$ . The inclusion of linear deterministic trends, however, has not been analysed in the literature before.

Let  $M$  be an  $n \times n$  nonsingular matrix given by  $[S'_k \ \alpha']$ , where the rows of the  $k \times n$  selection matrix  $S_k$  satisfy  $S_{i,k}C(1) \neq 0$  for all  $i \in \{1, \dots, k\}$ . Also, let  $N$  be an  $n \times n$  matrix equal to  $[0 \ \gamma]$ , while the  $n \times n$  matrix polynomials  $D(\lambda)$  and  $\Delta(\lambda)$  are

$$D(\lambda) := \begin{bmatrix} I_k & 0 \\ 0 & (1-\lambda)I_r \end{bmatrix}, \quad \Delta(\lambda) := \begin{bmatrix} (1-\lambda)I_k & 0 \\ 0 & I_r \end{bmatrix},$$

The main representation theorem of this thesis for a real valued vector time series with  $k$  common trends is the following:

**Theorem 2.1** *If and only if the  $n$  dimensional vector time series  $\{x_t\}_{t=1}^{\infty}$  (i) satisfies  $\Delta x_t = C(L) \cdot (\rho + \varepsilon_t)$ , where the  $n \times n$  matrix polynomial  $C(\lambda) = I_n + \sum_{j=1}^{\infty} C_j \lambda^j$ ,  $\{\varepsilon_t\}_{t=1}^{\infty}$  is a sequence of a purely nondeterministic and serially uncorrelated random variable,  $\varepsilon_t$ , with mean zero and positive definite covariance matrix  $\Sigma$ , and (ii) is cointegrated of order  $(1,1)$  with cointegrating rank  $r$ , then there exists a restricted vector autoregressive representation*

$$B(L)y_t = \theta + \eta_t. \quad (2.9)$$

Here,  $B(\lambda) := M[A^*(\lambda)M^{-1}D(\lambda) + N \cdot \lambda]$ ,  $B(0) = I_n$ , the  $n$  dimensional vector time series  $\{y_t\}_{t=1}^{\infty}$  is integrated of order zero,  $y_t := \Delta(L)Mx_t$ ,  $\theta := M\rho$ ,  $\eta_t := M\varepsilon_t$ ,  $E[\eta_t \eta'_t] = \Omega$  is positive definite, and the function  $\det[B(\lambda)] = 0$  has all solutions outside the unit circle. In addition,

$$A(\lambda) = M^{-1}B(\lambda)\Delta(\lambda)M, \quad (2.10)$$

and

$$C(\lambda) = M^{-1}D(\lambda)B(\lambda)^{-1}M. \quad (2.11)$$

**Proof** To show sufficiency, let us start from the vector error correction representation, which is known to exist under conditions (i) and (ii) (cf. Hylleberg and Mizon [42]). Premultiplying both sides of equation (2.3) by the nonsingular matrix  $M$ , we find that

$$MA^*(L)M^{-1}\Delta Mx_t = \theta - M\gamma z_{t-1} + \eta_t.$$

Making use of the relationships  $(1-\lambda)I_n = D(\lambda)\Delta(\lambda)$  and  $Ny_t = \gamma z_t$  we arrive at the restricted vector autoregressive representation for  $x_t$ . Since  $\eta_t = M\varepsilon_t$ , we find that  $E[\eta_t \eta'_t] = \Omega = M\Sigma M'$ . This matrix is evidently nonsingular by virtue of the fact that  $\Sigma$  and  $M$  are nonsingular matrices. Finally, the solutions to  $\det[B(\lambda)] = 0$  are outside the unit circle if the vector time series  $\{y_t\}_{t=1}^{\infty}$  is integrated of order zero. The latter is established by noting that  $y_t = [\Delta(S_k x_t)' \ z'_t]'$ .

Necessity is proven by first noting that the restricted vector autoregressive representation can be written as

$$M^{-1}B(L)\Delta(L)Mx_t = \rho + \varepsilon_t,$$

by premultiplying both sides in (2.9) by  $M^{-1}$ . From this expression we find that the matrix polynomial  $A(\lambda)$  in the unrestricted vector autoregressive representation is equal

to  $M^{-1}B(\lambda)\Delta(\lambda)M$ , where the rank of  $A(1)$  is equal to the rank of  $\Delta(1)$ . The latter matrix, of course, has rank  $r$ . Premultiplying both sides in the above equation by  $M^{-1}D(\lambda)F(\lambda)M$ , where  $F(\lambda) = B(\lambda)^{-1}$ , we find that

$$\Delta x_t = M^{-1}D(1)F(1)M\rho + M^{-1}D(L)F(L)M\varepsilon_t.$$

Hence,  $C(\lambda) = M^{-1}D(\lambda)F(\lambda)M$  is finite since  $F(\lambda)$  is finite and invertible, and the rank of  $C(1)$  is equal to the rank of  $D(1)$ , which is  $k$ . The latter property implies that  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with cointegrating rank  $r$  since  $\{y_t\}_{t=1}^{\infty}$  is integrated of order zero. Q.E.D.

In a sense, Theorem 2.1 summarizes all we need to know about the (reduced form) mathematical properties of a vector time series which is cointegrated of order (1,1) with cointegrating rank  $r$ . The matrix polynomial  $B(\lambda)$  captures the general ‘short run’ dynamics, whereas  $(\Delta(\lambda), D(\lambda))$  and  $M$  represent integration and cointegration, respectively (see Warne [98, page 20]). Furthermore, it provides a foundation for the Stock and Watson [95] strategy to testing for cointegration. To see this, note that  $S_k$  may be chosen so that its rows are orthogonal to the columns of the cointegrating matrix  $\alpha$ . Moreover, the restricted vector autoregression may be used as a convenient data generating process for testing linear rational expectations models when the time series of interest are cointegrated (cf. Baillie [4], Campbell and Shiller [14], Warne [98], and Chapter 4). However, for my purpose here, a more important result is that we have found a fundamental and simple mathematical connection to the vector moving average representation. Hence, the restricted vector autoregressive representation is very well suited for estimating a common trends model.<sup>3</sup>

## 2.2.2 Estimation of the Common Trends Parameters

Without loss of generality, let us suppose that the selection matrix,  $S_k$ , is given by  $[I_k \ 0]$ . This implies that  $y_t$  is equal to  $[\Delta x_{1,t} \cdots \Delta x_{k,t} \ z_{1,t} \cdots z_{r,t}]'$ .<sup>4</sup> From the relationship between  $A(\lambda)$  and  $B(\lambda)$  it follows that the latter matrix polynomial is never of an order greater than  $A(\lambda)$ . Unless all elements in the final  $r$  columns of  $A_p$  are zero,  $B(\lambda)$  is also of order  $p$ . Hence, I shall let  $B(\lambda) = I_n - \sum_{k=1}^p B_k \lambda^k$ . Furthermore, Theorem 2.1 establishes that the matrix  $C(1)$  is equal to  $M^{-1}D(1)F(1)M$ , where  $F(1)$  is the inverse of  $B(1)$ . It then follows that if  $M$ ,  $\theta$ ,  $\Omega$ , and  $B(1)$  were known, so would  $\rho$ ,  $\Sigma$ , and  $C(1)$  be.

The space spanned by the rows of  $\alpha'$  may be estimated and analysed by applying the maximum likelihood based methods developed in Johansen [44,45,48], and Johansen

---

<sup>3</sup>It may be noted that from a purely mathematical point of view it is *always* possible to consider a selection matrix  $S_k$  of the form  $S_k = [I_k \ 0]$  when the cointegrating vectors are known. The reason for this is that  $\text{rank}[\alpha] = r$  so that the components of  $x_t$  can be ordered to ensure that the last  $r$  columns of  $\alpha'$  is an invertible matrix. In fact, we can let  $\alpha' = [\alpha'_k \ I_r]$ , where  $\alpha'_k$  is an  $r \times k$  matrix. It is now easily established that  $\alpha'C(1) = 0$ .

<sup>4</sup>Several authors have actually applied the restricted autoregressive representation in equation (2.9) and this specific choice for  $S_k$  (cf. Campbell [13], Shapiro and Watson [86], and Walsh [97]). Intuitively, it seems natural to consider this choice for  $y_t$  when time series are cointegrated. Theorem 2.1 provides mathematical conditions for that intuition.

and Juselius [49]. Another possibility is to let these parameters be determined by the steady state of an appropriate economic theory (cf. King, Plosser, Stock and Watson [53] and Chapter 3). In either case, knowledge of these parameters suffices for the purpose of determining the matrices  $M$  and  $\Delta(\lambda)$ , needed to construct the vector time series  $\{y_t\}_{t=1}^{\infty}$ . Furthermore, consistent and asymptotically efficient estimates of the parameters  $\{\theta, B_1, \dots, B_p, \Omega\}$  may be obtained from, e.g., multivariate least squares or Gaussian maximum likelihood estimation of  $y_t$  on a constant and  $p$  lags (see, e.g., Baillie [4]).<sup>5</sup>

To summarize, we can calculate estimates of  $C(1)$ ,  $\Sigma$ , and  $\rho$  from estimates of the parameters in the restricted vector autoregressive representation and the exact relationships are:

$$\begin{aligned}\rho &= M^{-1}\theta, \\ \Sigma &= M^{-1}\Omega(M')^{-1}, \\ C(1) &= M^{-1}D(1)F(1)M.\end{aligned}\tag{2.12}$$

The next step is to calculate the matrix of common trends parameters,  $A$ , and the variance matrix for the innovations,  $\Phi$ . In order to identify the common trends,  $\tau_t$ , one may proceed along the route suggested by King, Plosser, Stock, and Watson [53] (see also, e.g., Johansen [44, Theorem 4.1] for an alternative approach). That is, when  $\{x_t\}_{t=1}^{\infty}$  has  $k$  common stochastic trends, we may write the matrix  $A$  as

$$A = A_0\pi,\tag{2.13}$$

where  $A_0$  is an  $n \times k$  matrix with known parameters, chosen so that  $\alpha' A_0 = 0$ , and  $\pi$  is a  $k \times k$  matrix of unknown parameters. Using the result that  $A\Phi A' = C(1)\Sigma C(1)'$  and letting  $\pi^* := \pi\Phi^{1/2}$ , we have that

$$A_0\pi^*\pi^{*'}A_0' = C(1)\Sigma C(1)'.\tag{2.14}$$

Since  $A_0$  has rank equal to  $k$ , we may rearrange the parameters of equation (2.14) in the following form

$$\pi^*\pi^{*'} = (A_0'A_0)^{-1}A_0'C(1)\Sigma C(1)'A_0(A_0'A_0)^{-1}.\tag{2.15}$$

The right hand side of equation (2.15) is a  $k \times k$  positive definite and symmetric matrix with known parameters. Accordingly,  $\pi^*\pi^{*'}$  is known.

We cannot, however, solve for  $\pi^*$  *uniquely* without making some additional assumptions. For the above system of equations no more than  $k \cdot (k + 1)/2$  parameters can be uniquely determined. Since  $\Phi$  is assumed to be diagonal, this implies that  $k \cdot (k - 1)/2$  parameters remain for computing  $\pi$  unless, for example, we choose  $\Phi = I_k$ ; then  $k(k + 1)/2$  parameters of  $\pi$  can be uniquely determined. In the first case, we may, e.g., suppose that  $\pi$  is lower triangular with unit diagonal elements. On the other hand, if  $\Phi = I_k$  is

---

<sup>5</sup>Note that the asymptotic properties of, e.g.,  $C(1)$  are not independent of how  $\alpha$  has been determined. An estimated  $\alpha$  matrix will always imply that the estimated  $C(1)$  matrix has a greater asymptotic covariance matrix than when  $\alpha$  is given by economic theory in the sense that the difference between the two covariance matrices is a positive definite matrix. More importantly,  $S_k = [I_k \ 0]$  need not be an appropriate choice for the selection matrix when  $\alpha$  is estimated since there is no guarantee that the estimated  $M$  matrix converges to a nonsingular matrix in probability. Instead, we can consider  $S_k = \alpha'_\perp$ , where  $\alpha'_\perp \alpha_\perp = 0$  and  $\alpha'_\perp \alpha_\perp = I_k$ . This selection matrix ensures that an estimate of  $M$ , based on appropriate assumptions (cf. Lütkepohl and Reimers [62]), converges in probability to a nonsingular matrix.

chosen, the diagonal elements of the lower triangular matrix  $\pi$  are free.<sup>6</sup> From a Choleski decomposition of  $\pi^*\pi'^$  we then obtain  $\pi$ ,  $\Phi$ , and  $A$ .

Other identification procedures such as a method of moments decomposition of  $\pi^*\pi'^$  may also be considered (cf. Bernanke [8]). It should be noted that although the Choleski decomposition of  $\pi$  indicates a recursive structure for the influence of  $\tau_t$  on  $x_t$ , the choice of  $A_0$  actually determines which components of  $\tau_t$  will have influence on  $x_t$  and which may not. Thus, unless one so desires,  $A$  need *not* represent any recursiveness for the common trends model. Finally, the  $k \times 1$  vector of drifts, associated with the mutually independent random walks,  $\tau_t$ , may be calculated from

$$\mu = (A'A)^{-1}A'C(1)M^{-1}\theta.$$

At this stage, it should be emphasized that this procedure for identifying the common trends parameters implies that the innovations to the common trends influence transient fluctuations in  $x_t$  as well as the growth path. To see this, note that  $\varphi_t = (A'A)^{-1}A'C(1)\varepsilon_t$ . Consequently, the covariance matrix between  $\varphi_t$  and  $\varepsilon_t$  is

$$E[\varphi_t\varepsilon_t'] = (A'A)^{-1}A'C(1)\Sigma.$$

Obviously, this matrix is nonzero since the columns of  $A$  cannot be orthogonal to the columns of  $C(1)$ . It is precisely this fact that allows us to study connections between growth and transitory fluctuations.

### 2.2.3 An Example

Suppose we are interested in examining the interactions between the logarithms of real output ( $\ln Y_t$ ), the price level ( $\ln P_t$ ), and the money stock ( $\ln M_t$ ). Furthermore, suppose we model  $x_t = [\ln Y_t \ln P_t \ln M_t]'$  as being cointegrated of order (1,1) with one cointegrating vector. Let  $\alpha' = [1 \ 1 \ -1]$ , so that the logarithm of the velocity of money is integrated of order zero. Accordingly, each series is nonstationary in levels.<sup>7</sup>

In this case, we can estimate the restricted vector autoregression in equation (2.9) with  $y_t = [\Delta \ln Y_t \ \Delta \ln P_t \ \ln V_t]'$ , where  $\ln V_t = (\ln Y_t + \ln P_t - \ln M_t)$ . From estimates of the  $B_k$  matrices we can then determine an estimate of  $C(1)$  as shown above.

To estimate the  $3 \times 2$  matrix of common trends parameters,  $A$ , we need to specify  $A_0$ . A choice suggested by economic theory is to let  $\tau_t$  include a real (technology) and a nominal (monetary policy) trend, and assume that the nominal trend does not influence

---

<sup>6</sup>Since the elements of  $\varphi_t$  can only be identified up to a linear (nonzero) transformation, it is irrelevant from a practical and theoretical point of view which specific nonzero numbers that are assigned to the diagonal of  $\Phi$ .

<sup>7</sup>Note, however, it is *not* necessary that each time series in a common trends model is nonstationary. For example, the time series model studied by Blanchard and Quah [10] can easily be fitted into a common trends framework. Since they examine real output and unemployment and model the former as first difference stationary and the latter as stationary, there is one cointegrating vector which assigns a nonzero coefficient to unemployment and a zero coefficient to real output. Then, the time series  $y_t$  is given by the first difference of real output and the level of unemployment. Furthermore, the matrix  $A$  is  $2 \times 1$  with a nonzero coefficient in the output equation and a zero coefficient in the unemployment equation. See also equation (1.12).

the long run growth path of  $\ln Y_t$ . The real trend, however, is *allowed* to influence growth in the nominal variables.

Given  $\alpha$ , the above discussion implies that

$$A_0 = \begin{bmatrix} a_{0,11} & 0 \\ a_{0,21} & a_{0,22} \\ a_{0,11} + a_{0,21} & a_{0,22} \end{bmatrix}.$$

It can be verified that the choice of  $a_{0,21}$  does *not* influence the matrix  $A$ , whereas  $a_{0,11}$  and  $a_{0,22}$  must be nonzero. Therefore, let  $a_{0,21} = 0$  and  $a_{0,11} = a_{0,22} = 1$ .<sup>8</sup> With  $\pi$  being lower triangular, the general form of  $A$  is then

$$A = \begin{bmatrix} \pi_{11} & 0 \\ \pi_{21} & \pi_{22} \\ \pi_{11} + \pi_{21} & \pi_{22} \end{bmatrix},$$

where  $\pi_{ij}$  denotes the  $(i, j)$ :th element of  $\pi$ . From  $A$  it can be seen that the first element of  $\varphi_t$  is the real trend, while the second element is the nominal trend. Furthermore, the matrix  $A$  satisfies three restrictions and thus has three free parameters, i.e.,  $a_{12} = 0$ ,  $a_{11} + a_{21} = a_{31}$ , and  $a_{22} = a_{32}$ .

It should be noted that if  $\pi_{ii}$  is obtained directly from  $\pi_{ii}^*$ , then the variance matrix of  $\varphi_t$  is given by  $I_2$ . On the other hand, if we let  $\pi_{ii} = 1$ , then  $\pi_{11}^{*2}$  is the variance of the innovation to the real trend, while  $\pi_{22}^{*2}$  is the variance of the innovation to the nominal trend. From a theoretical point of view, it is irrelevant which specific nonzero numbers we assign to the diagonal of  $\Phi$  since the common trends can only be identified up to a linear (nonzero) transformation.

## 2.3 Inversion and Identification

### 2.3.1 The Moving Average Parameters

The vector moving average representation in equation (2.1) is a natural starting point for analysing some dynamic properties of a vector time series  $\{x_t\}_{t=1}^{\infty}$  with  $k$  common trends. The central issues for performing impulse response analysis and forecast error variance decompositions are those of (a) calculating the sequence of matrices  $\{C_j\}_{j=1}^{\infty}$ , (b) identifying the innovations to the system, and (c) computing standard errors for the estimated parameters of these functions. Here, I shall focus on (a), whereas (b) and (c) are studied in sections 2.3.2 and 2.4, respectively.

From Theorem 2.1 we find that the  $C(\lambda)$  polynomial is equal to  $M^{-1}D(\lambda)F(\lambda)M$ , and  $F(\lambda)$  is the inverse of  $B(\lambda)$ . Since the latter polynomial is assumed to be of finite

---

<sup>8</sup>To verify this claim, let  $\phi$  be a  $k \times k$  matrix such that  $A = A_0\phi^{-1}\phi\pi = \tilde{A}_0\tilde{\pi}$ . Since  $\pi$  is lower triangular it follows that  $\phi$  must also be lower triangular for  $\tilde{\pi} := \phi\pi$  to be lower triangular. Standard matrix theory tells us that the inverse of a lower triangular matrix is lower triangular. Thus, the matrix  $\tilde{A}_0 := A_0\phi^{-1}$  can be constructed in any way we desire as long as  $\alpha'A_0 = 0$  and  $\phi$  is lower triangular. In the example, we may consider a lower triangular  $\phi$  matrix whose diagonal elements are given by  $\phi_{ii} = a_{0,ii}$  for  $i \in \{1, 2\}$ , thus the nonzero requirement. Also,  $\phi_{21} = a_{0,21}$  is permissible. If  $\pi$  is not lower triangular it becomes more difficult to determine what kind of  $\phi$  matrices that may be considered.

order  $p$ , it follows that  $F(\lambda)$  is not. Accordingly,  $C(\lambda)$  is also of infinite order. Letting  $F(\lambda) = I_n + \sum_{j=1}^{\infty} F_j \lambda^j$ , and using the above relationship, we find that

$$C(\lambda) = I_n + \sum_{j=1}^{\infty} M^{-1}(F_j - DF_{j-1})M\lambda^j, \quad (2.16)$$

where  $F_0 = I_n$  and the  $n \times n$  matrix  $D$  is defined from  $D(\lambda) = I_n - D \cdot \lambda$ . It may be noted that  $D = \Delta(1)$  is idempotent, a property I shall use below.

One algorithm for performing the inversion is obtained from using the fact that

$$F(\lambda)B(\lambda) = I_n + \sum_{j=1}^{\infty} (F_j - \sum_{i=0}^{j-1} F_i B_{j-i}) \cdot \lambda^j = I_n.$$

If this equality is to hold for any  $\lambda$ , it is evident that

$$F_j = \sum_{i=0}^{j-1} F_i B_{j-i}, \quad (2.17)$$

for all  $j \geq 1$ , where  $B_j = 0$  for all  $j \geq p + 1$ .

Alternatively, we can stack equation (2.9) into a first order system of the form

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \theta \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I_n & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & & I_n & 0 \end{bmatrix} \cdot \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

or

$$Y_t = \Theta + BY_{t-1} + N_t. \quad (2.18)$$

Since  $\det[B(\lambda)] = 0$  has all solutions outside the unit circle, it is clear that the eigenvalues of  $B$  are inside the unit circle. Accordingly,  $\lim_{s \rightarrow \infty} B^s = 0$  so that the solution to the system of stochastic difference equations in (2.18) is

$$Y_t = \sum_{j=0}^{\infty} B^j \Theta + \sum_{j=0}^{\infty} B^j N_{t-j}. \quad (2.19)$$

Defining the  $n \times np$  matrix  $J_p$  as  $[I_n \ 0 \ \cdots \ 0]$ , we find that  $y_t = J_p Y_t$ ,  $\Theta = J_p' \theta$ , and  $N_t = J_p' \eta_t$ . Hence, the solution to equation (2.9) in terms of current and past realizations of  $\eta_t$  can be written as

$$y_t = \sum_{j=0}^{\infty} J_p B^j J_p' \theta + \sum_{j=0}^{\infty} J_p B^j J_p' \eta_{t-j}, \quad (2.20)$$

where  $F(1) = \sum_{j=0}^{\infty} J_p B^j J_p'$ .<sup>9</sup> It is then obvious that an equivalent expression to that in equation (2.17) is

$$F_j = J_p B^j J_p'. \quad (2.21)$$

---

<sup>9</sup>This immediately follows from an inspection of equation (2.9). Inverting the  $B(\lambda)$  polynomial we find that  $y_t = B(1)^{-1}\theta + B(L)^{-1}\eta_t$ , where  $F(1) = B(1)^{-1}$ .

This latter relationship between  $F_j$  and the  $B_k$  matrices is simpler to work with when we derive asymptotic distributions for impulse response functions and forecast error variance decompositions; a matter to be discussed in section 2.4.

Hence, we find that the sequence of matrices  $\{C_j\}_{j=1}^{\infty}$  can be computed from the estimates of  $\{B_k\}_{k=1}^p$  and prior knowledge of  $M$ . In particular,

$$C_j = M^{-1}J_p B^j J_p' M - M^{-1}D J_p B^{j-1} J_p' M, \quad j = 1, 2, \dots \quad (2.22)$$

where  $C_0 = I_n$ .

### 2.3.2 Identification of Permanent and Transitory Innovations

In section 2.2.2 we identified  $k$  innovations, i.e., those which are associated with the common trends. In this subsection my objective is to be more specific about all innovations in the common trends model. In particular, it is important to be precise about identification in the sense that implications from impulse response functions and forecast error variance decompositions are fully consistent with the common trends model. Before we come to that, however, two definitions and some new notation is introduced to minimize ambiguities.

Let  $F$  be any  $n \times n$  nonsingular matrix such that  $F\Sigma F'$  is diagonal. The matrix  $R(1) = C(1)F^{-1}$  is called the total impact matrix. Also, let  $\nu_{it}$  be the  $i$ :th component of the vector  $F\varepsilon_t$ .

**Definition 2.1** *An innovation  $\nu_{it}$  is said to be permanent (transitory) if the  $i$ :th column of the total impact matrix is nonzero (zero).*

**Definition 2.2** *An  $n \times n$  matrix  $F$  is said to identify a common trends model if (i) the covariance matrix of  $F\varepsilon_t$  is diagonal with nonzero diagonal elements, and (ii) the total impact matrix is given by  $R(1) = [A \ 0]$ .*

From these two definitions it follows that if an  $n \times n$  matrix  $F$  identifies a common trends model, then the permanent innovations are those which are associated with the common trends.

Let the  $n \times n$  nonsingular matrix  $F$  be chosen so that (i) the permanent innovations are equal to  $\varphi_t$ , (ii) the permanent and the transitory innovations,  $\psi_t$ , are independent, and (iii) the transitory innovations are mutually independent. We then have that

$$\Delta x_t = \delta + C(L)\varepsilon_t = \delta + R(L)\nu_t, \quad (2.23)$$

where  $R(\lambda) = C(\lambda)F^{-1}$ ,  $\nu_t = F\varepsilon_t$ , and  $E[\nu_t\nu_t'] = \Gamma$ . The component  $R(L)\nu_t$  in equation (2.23) is called the impulse response function of  $\Delta x_t$ . The above three conditions imply that  $\Gamma$  is diagonal. Furthermore, this matrix can be partitioned into

$$\Gamma = \begin{bmatrix} \Phi & 0 \\ 0 & \Psi \end{bmatrix}, \quad (2.24)$$

where  $E[\psi_t\psi_t'] = \Psi = \text{diag}[\Psi_{11} \cdots \Psi_{rr}]$  is the variance matrix of the transitory innovations.

In order to derive a suitable matrix  $F$ , it may first be noted that

$$\nu_t = \begin{bmatrix} \varphi_t \\ \psi_t \end{bmatrix} = \begin{bmatrix} F_k \\ F_r \end{bmatrix} \varepsilon_t = F \varepsilon_t, \quad (2.25)$$

where  $F_k$  and  $F_r$  are  $k \times n$  and  $r \times n$  matrices, respectively. It has already been established that  $A\varphi_t = C(1)\varepsilon_t$  and that  $A$  as well as  $C(1)$  had rank equal to  $k$ . Hence, it is evident that the permanent innovations may be described by

$$\varphi_t = (A'A)^{-1}A'C(1)\varepsilon_t, \quad (2.26)$$

and, accordingly, the top  $k \times n$  matrix  $F_k$  in (2.25) is

$$F_k = (A'A)^{-1}A'C(1). \quad (2.27)$$

From section 2.2 we know that  $\Phi = F_k \Sigma F_k'$  is a diagonal matrix, which is implied by the assumption that the permanent innovations are mutually independent.

To find a matrix  $F_r$  which satisfies the conditions (ii)  $\varphi_t$  and  $\psi_t$  are independent, and (iii) the components of  $\psi_t$  are mutually independent, we may either make use of a Jordan decomposition of some suitable matrix or mathematically related schemes (cf. Stock and Watson [95]). I shall first consider condition (ii). Evaluating the covariance between the permanent and transitory innovations, we find that

$$E[\varphi_t \psi_t'] = (A'A)^{-1}A'C(1)\Sigma F_r'. \quad (2.28)$$

For this  $k \times r$  matrix to be zero, it seems natural to let  $F_r$  include  $\Sigma^{-1}$ . That allows us to focus on the matrix  $C(1)$ , which is known to have reduced rank. From linear algebra it is well known that there exists exactly  $r$  linearly independent vectors in  $\mathcal{R}^n$ , the  $n$  dimensional Euclidean space, which are orthogonal to the rows of  $C(1)$ . Letting  $F_r = H_r \Sigma^{-1}$ , we are therefore seeking an  $r \times n$  matrix  $H_r$  such that

$$C(1)H_r' = 0.$$

One possibility is to consider the space spanned by the columns of  $\gamma$ . From the properties of the  $A(1)$  and  $C(1)$  matrices, we have that  $C(1)\gamma = 0$ . Furthermore, from Theorem 2.1 we find that  $\gamma$  may be calculated from the estimated parameters  $\{B_k\}_{k=1}^p$  and the cointegrating matrix  $\alpha$ . In fact, the following relationship may be established

$$\gamma = M^{-1}B(1)\Delta(1)M\alpha(\alpha'\alpha)^{-1} = M^{-1}B(1)P_r, \quad (2.29)$$

where  $P_r$  is the  $n \times r$  matrix determined from  $D = [0 \ P_r]$ , i.e.,  $P_r = [0 \ I_r]'$ .<sup>10</sup> Premultiplying  $\gamma$  in equation (2.29) by  $C(1)$ , we find that

$$\begin{aligned} C(1)\gamma &= (M^{-1}D(1)F(1)M) \cdot (M^{-1}B(1)\Delta(1)M\alpha(\alpha'\alpha)^{-1}) \\ &= 0, \end{aligned}$$

---

<sup>10</sup>This is easily established by noting that  $M\alpha(\alpha'\alpha)^{-1} = [(\alpha'\alpha)^{-1}\alpha'S_k' \ I_r]'$ . Premultiplying by  $D$  we obtain  $P_r$ . Alternatively, from Theorem 2.1 we know that  $B(1) = M[A^*(1)M^{-1}D(1)+N]$ . Premultiplying by  $M^{-1}$  and postmultiplying by  $P_r$ , we find that  $M^{-1}B(1)P_r = NP_r$ . Since  $N = [0 \ \gamma]$ , it follows that  $NP_r = \gamma$ .

since  $D$  is idempotent, i.e.,  $D(1)\Delta(1) = (I_n - D)D = 0$ .

Let  $H_r = Q_r^{-1}\gamma'$ , where  $Q_r$  is an  $r \times r$  matrix. The covariance matrix for the transitory innovations is then given by

$$E[\psi_t \psi_t'] = Q_r^{-1} \gamma' \Sigma^{-1} \gamma (Q_r')^{-1} = \Psi. \quad (2.30)$$

In order to ensure that  $\Psi$  is compatible with the assumption of  $\psi_t$  being mutually independent,  $Q_r$  must be chosen such that  $\gamma' \Sigma^{-1} \gamma$  is diagonalized. In general, one may let the transitory innovations be identified in the sense that the innovations have an economic interpretation. Such identification involves, e.g., a method of moments (cf. Bernanke [8]) or a Choleski decomposition. However, if no specific economic interpretation of the transitory innovations is aimed at, then an ordinary eigen decomposition may also be considered. This decomposition implies that the diagonal of  $\Psi$  contains the eigenvalues of the positive definite  $r \times r$  matrix  $\gamma' \Sigma^{-1} \gamma$ , whereas the columnvectors of  $Q_r$  are the corresponding eigenvectors. Whichever decomposition we apply, the transitory innovations are determined from

$$\psi_t = Q_r^{-1} \gamma' \Sigma^{-1} \varepsilon_t. \quad (2.31)$$

Accordingly, the matrix  $F_r$  is given by  $Q_r^{-1} \gamma' \Sigma^{-1}$  so that the matrix  $F$  becomes

$$F = \begin{bmatrix} (A'A)^{-1} A' C(1) \\ Q_r^{-1} \gamma' \Sigma^{-1} \end{bmatrix}. \quad (2.32)$$

It may be noted that the  $k$  linearly independent rows of  $F_k$  are linearly independent to the  $r$  linearly independent rows of  $F_r$ . These properties imply that  $F$  is of full rank (cf. Theorem 3.19 in Magnus and Neudecker [63, page 56]).

We are now in a position to state the following important result concerning the properties of the matrix  $F$ .

**Theorem 2.2** *If the  $n$  dimensional vector time series  $\{x_t\}_{t=1}^{\infty}$  satisfies the assumptions in Theorem 2.1, then the  $n \times n$  nonsingular matrix  $F$  in equation (2.32) identifies a common trends model, i.e.,*

$$R(1) = C(1)F^{-1} = [A \ 0], \quad (2.33)$$

and  $F\Sigma F' = \Gamma$  is diagonal.

**Proof** Let us partition the inverse of  $F$  into

$$F^{-1} = \begin{bmatrix} F_k^{-1} & F_r^{-1} \end{bmatrix},$$

where  $F_k^{-1}$  and  $F_r^{-1}$  are  $n \times k$  and  $n \times r$  matrices, respectively. Postmultiplying  $F$  in equation (2.32) by this expression for  $F^{-1}$ , we obtain

$$FF^{-1} = \begin{bmatrix} (A'A)^{-1} A' C(1) F_k^{-1} & (A'A)^{-1} A' C(1) F_r^{-1} \\ Q_r^{-1} \gamma' \Sigma^{-1} F_k^{-1} & Q_r^{-1} \gamma' \Sigma^{-1} F_r^{-1} \end{bmatrix} = I_n.$$

Letting  $F_r^{-1} = \gamma(Q_r')^{-1} \Psi^{-1}$  and  $F_k^{-1} = \Sigma C(1)' A (A'A)^{-1} \Phi^{-1}$  we have found the inverse of  $F$ . Substituting for these relationships in equation (2.33), we have that

$$R(1) = [R(1)_k \ R(1)_r] = \begin{bmatrix} C(1) \Sigma C(1)' A (A'A)^{-1} \Phi^{-1} & C(1) \gamma (Q_r')^{-1} \Psi^{-1} \end{bmatrix}.$$

Clearly,  $R(1)_k = C(1) \Sigma C(1)' A (A'A)^{-1} \Phi^{-1} = A$  since  $C(1) \Sigma C(1)' = A \Phi A'$ , while  $R(1)_r = 0$  by virtue of the fact that  $C(1) \gamma = 0$ . Finally, it is obvious from the above analysis that  $F\Sigma F' = \Gamma$  is a diagonal matrix. Q.E.D.

## 2.4 Consistency and Asymptotic Normality

### 2.4.1 Asymptotic Properties of Impulse Response Functions

If the lag order of a covariance stationary vector autoregression is finite and known or if an upper bound for the order is known, results from Baillie [5,6], Schmidt [82,83], and others can be applied to obtain the asymptotic distribution of the vector moving average and impulse response parameters. Also, the case of unknown and possibly infinite lag order is examined by, e.g., Lütkepohl [58,59] and Lütkepohl and Poskitt [61].<sup>11</sup> Furthermore, the asymptotic distribution for the impulse responses in a cointegrated framework with Gaussian innovations is analysed in Lütkepohl and Reimers [62]. Unfortunately, neither of these approaches can *fully* be applied in the present setting. In particular, while the results in Lütkepohl and Reimers imply that some innovations will have permanent effect on some components of  $x_t$ , permanent and transitory innovations in the sense implied by the common trends model have generally *not* been identified.

Here, I shall concentrate on the case when an upper bound for the lag order as well as the cointegrating vectors are known. There are at least two reasons why the case of a known  $M$  matrix is of interest. First, economic theory can be made use of by letting the elements of the cointegrating vectors represent steady state parameters. At times, numerical values of these parameters can be assigned in much the same manner as is frequently the case in empirical studies on real business cycle models (cf. King, Plosser, Stock and Watson [53] and Chapter 3). Second, it is well known that prior information decreases the asymptotic covariances of the estimated parameters. The study by Runkle [80] suggests that the uncertainty of estimated impulse response functions in an unconstrained vector autoregression for covariance stationary time series is often disturbingly great. Although estimated cointegrating vectors impose a rank constraint on the parameters of the data generating process, it is unclear whether the effect on the forecasting uncertainty is substantial. Moreover, from Johansen [47], we find that the asymptotic local power of the likelihood ratio test for cointegration is generally, as one may suspect, low. If we wish to model data as being cointegrated and study a common trends model, it seems reasonable to suggest that economic theory should be consulted at least for determining steady states of the time series. For further discussions on this interesting topic, cf. King, Plosser, and Rebelo [52] and Runkle [80] with comments.

Without loss of generality, let  $\Gamma = I_n$ . Furthermore, suppose  $Q_r$  is lower triangular, i.e., the transitory innovations are computed from a Choleski decomposition of  $\gamma'\Sigma^{-1}\gamma$ .<sup>12</sup> Also, suppose that an upper bound for the lag order,  $p$ , is known. Let  $\text{vec}$  denote the column stacking operator for any matrix,  $\text{vech}$  the corresponding operator for symmetric and lower triangular matrices that only stacks elements on and below the diagonal. The Kronecker product is denoted by  $\otimes$ , the  $mn \times mn$  commutation matrix  $K_{mn}$  is defined such that for any  $m \times n$  matrix  $G$ ,  $K_{mn}\text{vec}(G) = \text{vec}(G')$ , and the  $m^2 \times m^2$  matrix  $N_m = \frac{1}{2}(I_{m^2} + K_{m,m})$ . The  $m^2 \times m(m+1)/2$  duplication matrix  $D_m$  is defined such that

---

<sup>11</sup>If it can be shown that a consistent and asymptotically normal estimator of  $B(1)$  exists when the lag order is unknown and possibly infinite and a finite order vector autoregression is fitted, it is straightforward to extend the theory below. I am currently examining the existence of such an estimator.

<sup>12</sup>Results similar to those obtained below hold with other  $Q_r$  and  $\pi$  matrices satisfying  $Q_r Q_r' = \gamma'\Sigma^{-1}\gamma$  and  $\pi\pi' = (A_0' A_0)^{-1} A_0' C(1)\Sigma C(1)' A_0 (A_0' A_0)^{-1}$ , respectively.

$D_m \text{vech}(G) = \text{vec}(G)$  for any symmetric  $m \times m$  matrix  $G$ . The Moore–Penrose inverse of  $D_m$  is given by  $D_m^+ = (D_m' D_m)^{-1} D_m'$ , while the  $m(m+1)/2 \times m^2$  elimination matrix  $L_m$  is defined such that for any  $m \times m$  matrix  $G$ ,  $\text{vech}(G) = L_m \text{vec}(G)$ . In addition, if  $G$  is lower triangular, then  $\text{vec}(G) = L_m' \text{vech}(G)$ ; see Henderson and Searle [40], Magnus and Neudecker [63,64,65], and Neudecker [67]. Finally, let  $\xrightarrow{d}$  denote convergence in distribution,  $T$  the sample size for the estimated parameters, which are denoted by a caret, while  $\mathcal{N}$  denotes the (multivariate) normal distribution.

Suppose that we shock  $\Delta x_t$  at  $t = t^*$  by a one standard deviation change in  $\nu_{t^*}$ . The dynamic responses in  $\Delta x_{t^*+s}$  are then given by

$$\text{resp}(\Delta x_{t^*+s}) = R_s, \quad (2.34)$$

where  $\text{resp}(\Delta x_{\text{inf}}) = \lim_{s \rightarrow \infty} \text{resp}(\Delta x_{t^*+s}) = 0$ . Similarly, the responses in the levels,  $x_{t^*+s}$ , are given by

$$\text{resp}(x_{t^*+s}) = \sum_{j=0}^s R_j, \quad (2.35)$$

where  $\text{resp}(x_{\text{inf}}) = \lim_{s \rightarrow \infty} \text{resp}(x_{t^*+s}) = R(1) = [A \ 0]$ . To estimate these impulse response functions we replace, e.g.,  $R_s$  with  $\hat{R}_s = \hat{C}_s \hat{F}^{-1}$ .

For the purpose of deriving asymptotic distributions of functions of estimated parameters a result from Serfling [85, Theorem 3.3.A] is employed. Let  $\phi \in \mathcal{R}^m$  be a vector of parameters and  $\hat{\phi}$  a consistent estimator of  $\phi$  such that

$$T^{1/2} (\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, V_\phi).$$

Suppose  $f(\phi)$  is a continuously differentiable function which maps  $\phi$  into  $\mathcal{R}^n$  and  $\partial f_i / \partial \phi' \neq 0$  at  $\phi$  for all  $i \in \{1, \dots, n\}$ . Then,

$$T^{1/2} (f(\hat{\phi}) - f(\phi)) \xrightarrow{d} \mathcal{N}(0, V_f),$$

where

$$V_f = \frac{\partial f}{\partial \phi'} V_\phi \frac{\partial f}{\partial \phi},$$

and  $(\partial f / \partial \phi)$  is the transpose of  $(\partial f / \partial \phi')$ . This well known result is vital in the following analysis. Hence, if  $\hat{\phi}$  is a consistent estimator of  $\phi$  and converges at the rate  $T^{1/2}$  to a joint asymptotic normal distribution, then  $f(\hat{\phi})$  has similar properties. Accordingly, only  $(\partial f / \partial \phi')$  has to be derived. It should perhaps be pointed out that instead of considering theoretical asymptotic distributions, one can employ resampling methods, such as the bootstrap and the jackknife to assess the uncertainty of the point estimates in question (cf. Efron [23]).

**Lemma 2.1** *If the lag order in equation (2.9) has a known upper bound  $p$ , the cointegrating vectors  $\alpha$  for an  $n$  dimensional vector time series  $\{x_i\}_{i=1}^\infty$  are known, and*

$$T^{1/2} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\omega} - \omega \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V),$$

where  $\beta = \text{vec}(J_p B)$ ,  $\omega = \text{vech}(\Omega)$ , and

$$V = \begin{bmatrix} V_\beta & 0 \\ 0 & V_\omega \end{bmatrix},$$

then

$$T^{1/2} \left( \text{vec}(\hat{F}_j) - \text{vec}(F_j) \right) \xrightarrow{d} \mathcal{N}(0, V_{F_j}), \quad (2.36)$$

for  $j = 1, 2, \dots$ , where

$$V_{F_j} = \frac{\partial \text{vec}(F_j)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(F_j)}{\partial \beta},$$

and

$$\frac{\partial \text{vec}(F_j)}{\partial \beta'} = \sum_{k=0}^{j-1} \left[ J_p(B')^{j-1-k} \otimes J_p B^k J_p' \right],$$

and  $F_k = J_p B^k J_p'$ ,

$$T^{1/2} \left( \sum_{i=1}^j \text{vec}(\hat{F}_i) - \sum_{i=1}^j \text{vec}(F_i) \right) \xrightarrow{d} \mathcal{N}(0, V_{\Sigma F_j}), \quad (2.37)$$

for  $j = 1, 2, \dots$ , where

$$V_{\Sigma F_j} = \left( \sum_{i=1}^j \frac{\partial \text{vec}(F_i)}{\partial \beta'} \right) V_\beta \left( \sum_{i=1}^j \frac{\partial \text{vec}(F_i)}{\partial \beta} \right),$$

and

$$T^{1/2} \left( \text{vec}(F(1)) - \text{vec}(F(1)) \right) \xrightarrow{d} \mathcal{N}(0, V_{F(1)}), \quad (2.38)$$

where

$$V_{F(1)} = \frac{\partial \text{vec}(F(1))}{\partial \beta'} V_\beta \frac{\partial \text{vec}(F(1))}{\partial \beta},$$

and

$$\frac{\partial \text{vec}(F(1))}{\partial \beta'} = [F(1)' E_p \otimes F(1)],$$

where  $E_p = [I_n \cdots I_n]$  is an  $n \times np$  matrix.

**Proof** The expression in equation (2.36) is proved by Baillie [5] using Lemma 1 in Schmidt [82], whereas Lütkepohl [59] proved (2.37) and (2.38). Q.E.D.

Regarding the asymptotic distributions for the parameters of the vector moving average representation, the following holds true:

**Lemma 2.2** *If the assumptions in Lemma 2.1 are satisfied, then*

$$T^{1/2} \left( \text{vec}(\hat{C}_j) - \text{vec}(C_j) \right) \xrightarrow{d} \mathcal{N}(0, V_{C_j}), \quad (2.39)$$

for  $j = 1, 2, \dots$ , where

$$V_{C_j} = \frac{\partial \text{vec}(C_j)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(C_j)}{\partial \beta},$$

and

$$\frac{\partial \text{vec}(C_j)}{\partial \beta'} = \sum_{k=0}^{j-1} [M' J_p (B')^{j-1-k} \otimes M^{-1} F_k] - \sum_{k=0}^{j-2} [M' J_p (B')^{j-2-k} \otimes M^{-1} D F_k],$$

where the second term on the right hand side is zero for  $j = 1$ , and

$$T^{1/2} \left( \sum_{i=1}^j \text{vec}(\hat{C}_i) - \sum_{i=1}^j \text{vec}(C_i) \right) \xrightarrow{d} \mathcal{N}(0, V_{\Sigma C_j}), \quad (2.40)$$

for  $j = 1, 2, \dots$ , where

$$V_{\Sigma C_j} = \left( \sum_{i=1}^j \frac{\partial \text{vec}(C_i)}{\partial \beta'} \right) V_{\beta} \left( \sum_{i=1}^j \frac{\partial \text{vec}(C_i)}{\partial \beta} \right),$$

and, finally,

$$T^{1/2} \left( \text{vec}(\hat{C}(1)) - \text{vec}(C(1)) \right) \xrightarrow{d} \mathcal{N}(0, V_{C(1)}), \quad (2.41)$$

where

$$V_{C(1)} = \frac{\partial \text{vec}(C(1))}{\partial \beta'} V_{\beta} \frac{\partial \text{vec}(C(1))}{\partial \beta},$$

and

$$\frac{\partial \text{vec}(C(1))}{\partial \beta'} = (M' \otimes M^{-1} D(1)) \frac{\partial \text{vec}(F(1))}{\partial \beta'}.$$

**Proof** To show (2.39), we make use of equation (2.22). The first differential is equal to:

$$dC_j = \sum_{k=0}^{j-1} M^{-1} J_p B^k dBB^{j-1-k} J_p' M - \sum_{k=0}^{j-2} M^{-1} D J_p B^k dBB^{j-2-k} J_p' M,$$

where the second term on the right hand side is zero for  $j = 1$  since  $dF_0 = 0$ . Noting that  $dB = J_p' db$ , where  $b = [B_1 \cdots B_p]$  and making use of the vec operator, the desired result is obtained. Expressions (2.40) and (2.41) follow from (2.39) in the same way as (2.37) and (2.38) follow from (2.36). Q.E.D.

We are now ready to state the main results regarding analytical expressions for the asymptotic distributions of the impulse response functions.

**Theorem 2.3** *If the assumptions in Lemma 2.1 are satisfied and the matrix  $F$  is given by equation (2.32), with  $\pi$  and  $Q_r$  being lower triangular matrices, then*

$$T^{1/2} \left( \text{vec}(\hat{R}_j) - \text{vec}(R_j) \right) \xrightarrow{d} \mathcal{N}(0, V_{R_j}), \quad (2.42)$$

for  $j = 0, 1, 2, \dots$ , where

$$V_{R_j} = \frac{\partial \text{vec}(R_j)}{\partial \beta'} V_{\beta} \frac{\partial \text{vec}(R_j)}{\partial \beta} + \frac{\partial \text{vec}(R_j)}{\partial \omega'} V_{\omega} \frac{\partial \text{vec}(R_j)}{\partial \omega},$$

and

$$\frac{\partial \text{vec}(R_j)}{\partial \beta'} = [(F^{-1})' \otimes I_n] \frac{\partial \text{vec}(C_j)}{\partial \beta'} + [I_n \otimes C_j] \left[ \frac{\partial \text{vec}(F_k^{-1})}{\partial \beta'} / \frac{\partial \text{vec}(F_r^{-1})}{\partial \beta'} \right],$$

with  $C_0 = I_n$  so that  $dC_0 = 0$ , while

$$\frac{\partial \text{vec}(R_j)}{\partial \omega'} = [I_n \otimes C_j] \left[ \frac{\partial \text{vec}(F_k^{-1})}{\partial \omega'} / \frac{\partial \text{vec}(F_r^{-1})}{\partial \omega'} \right],$$

Furthermore,

$$\begin{aligned} \frac{\partial \text{vec}(F_k^{-1})}{\partial \beta'} &= K_{nk} [M^{-1} \Omega F(1)' E_p \otimes (A' A)^{-1} A' M^{-1} D(1) F(1)] - \\ &\quad [\pi^{-1} \otimes \Sigma C(1)' A (A' A)^{-1}] K_{kk} L'_k \{L_k N_k (\pi \otimes I_k) L'_k\}^{-1} D_k^+ \times \\ &\quad [(A'_0 A_0)^{-1} A'_0 C(1) M^{-1} \Omega F(1)' E_p \otimes (A'_0 A_0)^{-1} A'_0 M^{-1} D(1) F(1)], \\ \frac{\partial \text{vec}(F_r^{-1})}{\partial \beta'} &= -[Q_r^{-1} P'_r E_p \otimes M^{-1}] + [Q_r^{-1} \otimes \gamma(Q'_r)^{-1}] K_{rr} \times \\ &\quad L'_r \{L_r N_r (Q_r \otimes I_r) L'_r\}^{-1} D_r^+ [P'_r E_p \otimes \gamma' M' \Omega^{-1}], \end{aligned}$$

whereas

$$\begin{aligned} \frac{\partial \text{vec}(F_k^{-1})}{\partial \omega'} &= [(A' A)^{-1} A' C(1) M^{-1} \otimes M^{-1}] D_n - \frac{1}{2} [\pi^{-1} \otimes \Sigma C(1)' A (A' A)^{-1}] K_{kk} \times \\ &\quad L'_k \{L_k N_k (\pi \otimes I_k) L'_k\}^{-1} D_k^+ [(A'_0 A_0)^{-1} A'_0 C(1) M^{-1} \\ &\quad \otimes (A'_0 A_0)^{-1} A'_0 C(1) M^{-1}] D_n, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \text{vec}(F_r^{-1})}{\partial \omega'} &= \frac{1}{2} [Q_r^{-1} \otimes \gamma(Q'_r)^{-1}] K_{rr} L'_r \{L_r N_r (Q_r \otimes I_r) L'_r\}^{-1} D_r^+ \times \\ &\quad [\gamma' M' \Omega^{-1} \otimes \gamma' M' \Omega^{-1}] D_n. \end{aligned}$$

The asymptotic distributions for the accumulated response functions are given by

$$T^{1/2} \left( \sum_{i=0}^j \text{vec}(\hat{R}_i) - \sum_{i=0}^j \text{vec}(R_i) \right) \xrightarrow{d} \mathcal{N}(0, V_{\Sigma R_j}), \quad (2.43)$$

for  $j = 0, 1, 2, \dots$ , where

$$V_{\Sigma R_j} = \left( \sum_{i=0}^j \frac{\partial \text{vec}(R_i)}{\partial \beta'} \right) V_\beta \left( \sum_{i=0}^j \frac{\partial \text{vec}(R_i)}{\partial \beta} \right) + \left( \sum_{i=0}^j \frac{\partial \text{vec}(R_i)}{\partial \omega'} \right) V_\omega \left( \sum_{i=0}^j \frac{\partial \text{vec}(R_i)}{\partial \omega} \right),$$

and finally,

$$T^{1/2} (\text{vec}(\hat{A}) - \text{vec}(A)) \xrightarrow{d} \mathcal{N}(0, V_A), \quad (2.44)$$

where

$$V_A = \frac{\partial \text{vec}(A)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(A)}{\partial \beta} + \frac{\partial \text{vec}(A)}{\partial \omega'} V_\omega \frac{\partial \text{vec}(A)}{\partial \omega},$$

while the remaining matrices of partial derivatives are given by

$$\frac{\partial \text{vec}(A)}{\partial \beta'} = [I_k \otimes A_0] L'_k \frac{\partial \text{vech}(\pi)}{\partial \beta'}$$

$$\begin{aligned} \partial \text{vech}(\pi) / \partial \beta' &= \{L_k N_k(\pi \otimes I_k) L_k'\}^{-1} D_k^+ \times \\ &[(A_0' A_0)^{-1} A_0 C(1) M^{-1} \Omega F(1)' E_p \otimes (A_0' A_0)^{-1} A_0' M^{-1} D(1) F(1)], \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \text{vec}(A)}{\partial \omega'} &= [I_k \otimes A_0] L_k' \frac{\partial \text{vech}(\pi)}{\partial \omega'}, \\ \partial \text{vech}(\pi) / \partial \omega' &= \frac{1}{2} \{L_k N_k(\pi \otimes I_k) L_k'\}^{-1} D_k^+ \times \\ &[(A_0' A_0)^{-1} A_0' C(1) M^{-1} \otimes (A_0' A_0)^{-1} A_0' C(1) M^{-1}] D_n. \end{aligned}$$

**Proof** In order to derive the matrices in equation (2.42), note that the following relationships hold:  $R_j = C_j F^{-1}$ ,  $F^{-1} = [\Sigma C(1)' A(A'A)^{-1} \gamma(Q_r')^{-1}]$ ,  $\Sigma = M^{-1} \Omega(M')^{-1}$ ,  $\gamma = M^{-1}(I_n - J_p B E_p') P_r$ ,  $A = A_0 \pi$ ,  $Q_r Q_r' = \gamma' \Sigma^{-1} \gamma$ , and

$$\pi \pi' = (A_0' A_0)^{-1} A_0' C(1) \Sigma C(1)' A_0 (A_0' A_0)^{-1}.$$

Then, taking  $\text{vec}$ 's of the first differential of  $R_j$ , we find that

$$d \text{vec}(R_j) = [(F^{-1})' \otimes I_n] d \text{vec}(C_j) + [I_n \otimes C_j] \begin{bmatrix} d \text{vec}(F_k^{-1}) \\ d \text{vec}(F_r^{-1}) \end{bmatrix}. \quad (2.45)$$

Before we consider  $d \text{vec}(F_k^{-1})$ , note that

$$A(A'A)^{-1} = A_0 \pi (\pi' A_0' A_0 \pi)^{-1} = A_0 (A_0' A_0)^{-1} (\pi')^{-1}.$$

Since  $A_0$  is known we only need to consider  $\pi$  when deriving the first differential of  $A(A'A)^{-1}$ .

The first differential of  $F_k^{-1}$  is equal to

$$\begin{aligned} dF_k^{-1} &= (d\Sigma) C(1)' A(A'A)^{-1} + \Sigma (dC(1)') A(A'A)^{-1} \\ &\quad - \Sigma C(1)' A_0 (A_0' A_0)^{-1} (\pi')^{-1} (d\pi') (\pi')^{-1}. \end{aligned}$$

Taking  $\text{vec}$ 's and using the commutation matrix, we have that

$$\begin{aligned} d \text{vec}(F_k^{-1}) &= [(A'A)^{-1} A' C(1) \otimes I_n] d \text{vec}(\Sigma) + [(A'A)^{-1} A' \otimes \Sigma] K_{nn} d \text{vec}(C(1)) \\ &\quad - [\pi^{-1} \otimes \Sigma C(1)' A(A'A)^{-1}] K_{kk} d \text{vec}(\pi). \end{aligned} \quad (2.46)$$

From the function mapping the parameters of  $\Omega$  into  $\Sigma$ , we get

$$d \text{vec}(\Sigma) = [M^{-1} \otimes M^{-1}] d \text{vec}(\Omega) = [M^{-1} \otimes M^{-1}] D_n d\omega. \quad (2.47)$$

Furthermore, we know from Lemma 2.1 and 2.2 above that

$$d \text{vec}(C(1)) = [M' F(1)' E_p \otimes M^{-1} D(1) F(1)] d\beta. \quad (2.48)$$

Next, the matrix  $\pi$  is lower triangular and obtained from a symmetric matrix. Lemma 1 in Lütkepohl [60] implies that

$$d \text{vech}(\pi) = \{2L_k N_k(\pi \otimes I_k) L_k'\}^{-1} d \text{vech}(\pi \pi').$$

From the properties of the elimination and duplication matrices, we then find that

$$d\text{vec}(\pi) = L'_k \{2L_k N_k (\pi \otimes I_k) L'_k\}^{-1} D_k^+ d\text{vec}(\pi\pi'). \quad (2.49)$$

The first differential of  $\pi\pi'$  is

$$\begin{aligned} d(\pi\pi') &= (A'_0 A_0)^{-1} A'_0 (dC(1)) \Sigma C(1)' A_0 (A'_0 A_0)^{-1} + \\ &\quad (A'_0 A_0)^{-1} A'_0 C(1) \Sigma (dC(1)') A_0 (A'_0 A_0)^{-1} + \\ &\quad (A'_0 A_0)^{-1} A'_0 C(1) (d\Sigma) C(1)' A_0 (A'_0 A_0)^{-1}. \end{aligned}$$

Taking vec's and making use of the commutation and  $N$  matrices, we find that

$$\begin{aligned} d\text{vec}(\pi\pi') &= 2N_k [(A'_0 A_0)^{-1} A'_0 C(1) \Sigma \otimes (A'_0 A_0)^{-1} A'_0] d\text{vec}(C(1)) + \\ &\quad [(A'_0 A_0)^{-1} A'_0 C(1) \otimes (A'_0 A_0)^{-1} A'_0 C(1)] d\text{vec}(\Sigma), \end{aligned} \quad (2.50)$$

by employing Lemma 4 in Magnus and Neudecker [65]. Substituting for the expressions in equation (2.47)–(2.50) into (2.46), using the relationship between  $\Sigma$  and  $\Omega$ , and the fact that  $D_k^+ N_k = D_k^+$  (Magnus and Neudecker [63, Theorem 3.12]), the matrices of partial derivatives of  $\text{vec}(F_k^{-1})$  with respect to  $\beta$  and  $\omega$  follow directly.

The first differential of  $F_r^{-1}$  is

$$dF_r^{-1} = (d\gamma)(Q'_r)^{-1} - \gamma(Q'_r)^{-1} (dQ'_r)(Q'_r)^{-1}.$$

Taking vec's we get

$$d\text{vec}(F_r^{-1}) = [Q_r^{-1} \otimes I_n] d\text{vec}(\gamma) - [Q_r^{-1} \otimes \gamma(Q'_r)^{-1}] K_{rr} d\text{vec}(Q_r). \quad (2.51)$$

Next, the first differential of  $\gamma$  is

$$d\gamma = -M^{-1} J_p (dB) E'_p P_r.$$

Employing the vec operator we obtain

$$d\text{vec}(\gamma) = -[P'_r E_p \otimes M^{-1} J_p] d\text{vec}(B) = -[P'_r E_p \otimes M^{-1}] d\beta, \quad (2.52)$$

since  $d\text{vec}(B) = [I_{np} \otimes J'_p] d\beta$  and  $J_p J'_p = I_n$ .

With  $Q_r$  being lower triangular, Lütkepohl [60, Lemma 1] provides us with

$$d\text{vec}(Q_r) = L'_r \{2L_r N_r (Q_r \otimes I_r) L'_r\}^{-1} D_r^+ d\text{vec}(Q_r Q'_r). \quad (2.53)$$

The first differential of  $Q_r Q'_r$  is

$$d(Q_r Q'_r) = (d\gamma') \Sigma^{-1} \gamma + \gamma' \Sigma^{-1} (d\gamma) - \gamma' \Sigma^{-1} (d\Sigma) \Sigma^{-1} \gamma.$$

Taking vec's, we get

$$d\text{vec}(Q_r Q'_r) = 2N_r [I_r \otimes \gamma' \Sigma^{-1}] d\text{vec}(\gamma) - [\gamma' \Sigma^{-1} \otimes \gamma' \Sigma^{-1}] d\text{vec}(\Sigma). \quad (2.54)$$

Substituting for the expressions in equations (2.47) and (2.52)–(2.54) into (2.51) and making use of the property  $D_r^+ N_r = D_r^+$ , the matrices with partial derivatives of  $\text{vec}(F_r^{-1})$

with respect to  $\beta$  and  $\omega$  are immediate. Finally, equations (2.43) and (2.44) follow directly from (2.42), the above expressions, and the fact that  $d\text{vec}(A) = [I_k \otimes A_0]d\text{vec}(\pi)$ . Q.E.D.

A few remarks about the asymptotic covariance matrices for consistent and asymptotically normal estimators of  $\beta$  and  $\omega$  may be of interest.<sup>13</sup> Here, I shall only consider the multivariate least squares and Gaussian maximum likelihood estimators of the parameters. First, it is well known that  $\hat{\beta}$  and  $\hat{\omega}$  are asymptotically independent under standard assumptions. Moreover, if  $\theta = 0$ , then  $V_\beta = (\mathcal{I}_Y^{-1} \otimes \Omega)$ , where  $\mathcal{I}_Y := p \lim T^{-1} \sum_{t=1}^T Y_{t-1} Y_{t-1}'$  is assumed to be nonsingular (cf. Warne [99]). If a constant is included in the restricted vector autoregression, the expression for  $V_\beta$  looks similar to that above. In fact, letting  $Y_t^* := [1 \ Y_t]'$ , the  $np \times (np + 1)$  matrix  $G = [0 \ I_{np}]$ , and  $\mathcal{I}_{Y^*} := p \lim T^{-1} \sum_{t=1}^T Y_{t-1}^* Y_{t-1}^{*'}$ , we find that  $V_\beta = (G \mathcal{I}_{Y^*}^{-1} G' \otimes \Omega)$ . Second, if  $\eta_t$  is i.i.d. Gaussian, from the inverse of the information matrix we find that  $V_\omega = 2D_n^+(\Omega \otimes \Omega)D_n^{+'}$  (cf. Magnus and Neudecker [65, section 11], or Warne [99]).

Furthermore, it can be seen that  $Q_r$  does not influence the upper left  $nk \times nk$  submatrices of  $V_{R_j}$  and  $V_{\Sigma R_j}$ . The diagonal elements of these matrices are the asymptotic variances of the responses in  $\Delta x_t$  and  $x_t$ , respectively, from a one standard deviation impulse to the permanent innovations. Similarly,  $\pi$  does not appear in the lower right  $nr \times nr$  submatrices of these covariance matrices. Hence, identification of the permanent and the transitory innovations can be said to be asymptotically independent. This is clearly a desirable property, whether we are primarily concerned with the permanent innovations or not.

Moreover, it should be noted that some of the asymptotic variances may be zero. For example, if an element of the  $k$ :th column of  $A_0$  is constrained to zero, then the corresponding elements of  $\hat{A}$  and  $A$  are also zero. Although this is really not troublesome from a theoretical point of view, in applied work one has to be cautious to this possibility when, e.g.,  $t$ -ratios are calculated (cf. Lütkepohl [59]).

Before we turn our attention to forecast error variance decompositions, the asymptotic properties of the drifts  $\delta$  and  $\mu$  are as follows:

**Corollary 2.1** *If in addition to the assumptions in Theorem 2.3*

$$T^{1/2} \begin{bmatrix} \hat{\beta}^* - \beta^* \\ \hat{\omega} - \omega \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V^*),$$

where  $\beta^* = \text{vec}([\theta \ J_p B])$ , and

$$V^* = \begin{bmatrix} V_{\beta^*} & 0 \\ 0 & V_\omega \end{bmatrix},$$

then

$$T^{1/2} (\hat{\delta} - \delta) \xrightarrow{d} \mathcal{N}(0, V_\delta), \quad (2.55)$$

where

$$V_\delta = \frac{\partial \delta}{\partial \beta^{*'}} V_{\beta^*} \frac{\partial \delta}{\partial \beta^*},$$

---

<sup>13</sup>For a general treatment, see Dunsmuir and Hannan [22].

and

$$\frac{\partial \delta}{\partial \beta^{*'}} = \left[ C(1)M^{-1} \quad (\theta'F(1)'E_p \otimes M^{-1}D(1)F(1)) \right],$$

while the asymptotic distribution of  $\hat{\mu}$  is

$$T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, V_\mu), \quad (2.56)$$

where

$$V_\mu = \frac{\partial \mu}{\partial \beta^{*'}} V_{\beta^*} \frac{\partial \mu}{\partial \beta^*} + \frac{\partial \mu}{\partial \omega'} V_\omega \frac{\partial \mu}{\partial \omega},$$

while

$$\frac{\partial \mu}{\partial \beta^{*'}} = \left[ (A'A)^{-1}A'C(1)M^{-1} \quad (\theta' \otimes \pi^{-1})\Xi(F(1)'E_p \otimes (A'_0A_0)^{-1}A'_0M^{-1}D(1)F(1)) \right],$$

$$\frac{\partial \mu}{\partial \omega'} = -[\theta'(M')^{-1}C(1)'A(A'A)^{-1} \otimes \pi^{-1}]L'_k \frac{\partial \text{vech}(\pi)}{\partial \omega'},$$

and

$$\Xi := I_{nk} - [(M')^{-1}C(1)'A(A'A)^{-1} \otimes I_k]L'_k \{L_k N_k(\pi \otimes I_k)L'_k\}^{-1} D_k^+ \times [(A'_0A_0)^{-1}A'_0C(1)M^{-1}\Omega \otimes I_k].$$

**Proof** We know that  $\delta = C(1)M^{-1}\theta$  and  $\mu = \pi^{-1}(A'_0A_0)^{-1}A'_0\delta$ , where  $\pi^{-1}(A'_0A_0)^{-1}A'_0 = (A'A)^{-1}A'$ . The first differential of  $\delta$  is

$$\begin{aligned} d\delta &= (dC(1))M^{-1}\theta + C(1)M^{-1}(d\theta) \\ &= [\theta'(M')^{-1} \otimes I_n]d\text{vec}(C(1)) + C(1)M^{-1}(d\theta). \end{aligned} \quad (2.57)$$

From Lemma 2.2 and the definition of  $\beta^*$  it follows that

$$d\delta = \left[ C(1)M^{-1} \quad (\theta'F(1)'E_p \otimes M^{-1}D(1)F(1)) \right] d\beta^*, \quad (2.58)$$

from which the matrix of partial derivatives of  $\delta$  with respect to  $\beta^*$  follows trivially.

The first differential of  $\mu$  is

$$\begin{aligned} d\mu &= -\pi^{-1}(d\pi)(A'A)^{-1}A'\delta + (A'A)^{-1}A'(d\delta) \\ &= -[\delta'A(A'A)^{-1} \otimes \pi^{-1}]d\text{vec}(\pi) + (A'A)^{-1}A'(d\delta). \end{aligned} \quad (2.59)$$

Substituting equations (2.49) and (2.50) for  $d\text{vec}(\pi)$ , equation (2.58) for  $d\delta$ , and rearranging terms the claimed result follows. Q.E.D.

In reference to the discussion on consistent and asymptotically normal estimators of  $\beta$  and  $\omega$ , it may be noted that  $V_{\beta^*} = (Z_Y^{-1} \otimes \Omega)$  when the restricted vector autoregression is estimated by multivariate least squares or Gaussian maximum likelihood.

## 2.4.2 Derivation of the Forecast Error Variance Expressions

In this subsection, I shall derive expressions of the  $s$  steps ahead forecast error covariance matrices of  $\Delta x$  and  $x$ . From these matrices formulae of the variance decompositions are easy to compute. It should be noted that the  $s$  steps ahead forecast error covariance matrix of  $\Delta x$  is a version of that stated for covariance stationary time series in, e.g., Reinsel [76] and Yamamoto [104], while an  $s$  steps ahead forecast error covariance matrix of  $x$  is derived in Reinsel and Lewis [77]. Finally, the long run variance decomposition of  $x$  is of particular interest in the common trends model. Since the  $s$  steps ahead forecast error covariance matrix of  $x$  is not finite when  $s \rightarrow \infty$ , transformations of the covariance matrix has to be considered. Furthermore, the long run variance decomposition of  $x$  must be numerically invariant with respect to any such transformation.

From equation (2.23) we find that

$$\begin{aligned}\Delta x_{t+s} &= \delta + \sum_{j=1}^{\infty} R_{j-1} \nu_{t+s+1-j} \\ &= \delta + \sum_{j=1}^s R_{j-1} \nu_{t+s+1-j} + \sum_{j=1}^{\infty} R_{s-1+j} \nu_{t+1-j}.\end{aligned}\tag{2.60}$$

The  $s$  steps ahead forecast of  $\Delta x_{t+s}$  based on information available at  $t$  is

$$E[\Delta x_{t+s} | \nu_t, \nu_{t-1}, \dots] = \delta + \sum_{j=1}^{\infty} R_{s-1+j} \nu_{t+1-j}.$$

Hence, the  $s$  steps ahead forecast error is

$$\epsilon_{t+s} = \sum_{j=1}^s R_{j-1} \nu_{t+s+1-j}$$

so the  $s$  steps ahead forecast error covariance matrix is given by

$$V_s := E[\epsilon_{t+s} \epsilon'_{t+s}] = \sum_{j=1}^s R_{j-1} R'_{j-1}.\tag{2.61}$$

The  $i$ :th diagonal element of  $V_s$  is the  $s$  steps ahead forecast error variance of  $\Delta x_i$ ,  $i = 1, \dots, n$ . Then

$$V_{i,s} := e'_i V_s e_i = \sum_{j=1}^s e'_i R_{j-1} R'_{j-1} e_i.\tag{2.62}$$

We can now decompose the expression on the right hand side of equation (2.62) into  $n$  terms. That is,

$$e'_i R_{j-1} = \left[ e'_i R_{j-1} e_1 \quad e'_i R_{j-1} e_2 \quad \dots \quad e'_i R_{j-1} e_n \right],$$

where  $e'_i R_{j-1} e_l$  is the  $(i, l)$ :th element of  $R_{j-1}$ . Hence,

$$V_{i,s} = \sum_{j=1}^s \sum_{l=1}^n (e'_i R_{j-1} e_l)^2.\tag{2.63}$$

Equations (2.62) and (2.63) may now be employed to construct the  $s$  steps ahead forecast error variance decomposition for  $x_i$ .

Moving  $x_{t+s-1}$  in equation (2.60) to the right hand side and recursively substituting for  $x_{t+s-1}, \dots, x_{t+1}$ , we obtain

$$\begin{aligned} x_{t+s} &= x_t + \delta s + \sum_{m=1}^s \sum_{j=1}^{\infty} R_{j-1} \nu_{t+m+1-j} \\ &= x_t + \delta s + \sum_{m=1}^s \sum_{j=1}^m R_{j-1} \nu_{t+m+1-j} + \sum_{m=1}^s \sum_{j=m+1}^{\infty} R_{j-1} \nu_{t+m+1-j}. \end{aligned} \quad (2.64)$$

The  $s$  steps ahead forecast of  $x_{t+s}$  based on information at  $t$  is

$$E[x_{t+s} | \nu_t, \nu_{t-1}, \dots] = x_t + \delta s + \sum_{m=1}^s \sum_{j=m+1}^{\infty} R_{j-1} \nu_{t+m+1-j}.$$

Accordingly, the  $s$  steps ahead forecast error for  $x_{t+s}$  is given by

$$\begin{aligned} \epsilon_{t+s}^* &= \sum_{m=1}^s \sum_{j=1}^m R_{j-1} \nu_{t+m+1-j} \\ &= \sum_{m=1}^s \left( \sum_{j=1}^{s+1-m} R_{j-1} \right) \nu_{t+m} \\ &= \sum_{m=1}^s \epsilon_{t+m}^*. \end{aligned}$$

Hence, the stochastic process  $\{\epsilon_{t+s}^*\}_{s=1}^{\infty}$  has a unit root and the forecast error variance will therefore explode linearly in  $s$ .

The  $s$  steps ahead forecast error covariance matrix is

$$\begin{aligned} V_s^* &= \sum_{m=1}^s \left( \sum_{j=1}^{s+1-m} R_{j-1} \right) \left( \sum_{j=1}^{s+1-m} R_{j-1}' \right) \\ &= \sum_{m=1}^s \left( \sum_{j=1}^m R_{j-1} \right) \left( \sum_{j=1}^m R_{j-1}' \right) \\ &= \sum_{m=1}^s R_m^* R_m^{*'} \end{aligned} \quad (2.65)$$

It should be noted that the second equality follows from reversing the order of summation. For  $s$  finite, the variance decomposition  $v_{ii,s}^*$  follows directly from equation (2.65) and the analysis on  $V_s$  above. However, the long run case deserves special attention since  $\lim_{s \rightarrow \infty} V_s^*$  is not finite.

If we collect the products  $R_{j-1} R_{j-1}'$  into one group and all other products into a second group, we find that

$$V_s^* = \sum_{j=1}^s (s+1-j) R_{j-1} R_{j-1}' + \sum_{m=1}^{s-1} \sum_{j=1}^m (s-m) (R_{j-1} R_m' + R_m R_{j-1}'). \quad (2.66)$$

Let  $\bar{V}_s := V_s^*/s$ . As  $s \rightarrow \infty$ , we get

$$\begin{aligned} \bar{V}_{\text{inf}} = \lim_{s \rightarrow \infty} \bar{V}_s &= \sum_{j=1}^{\infty} R_{j-1} R_{j-1}' + \sum_{m=1}^{\infty} \sum_{j=1}^m (R_{j-1} R_m' + R_m R_{j-1}') \\ &= \left( \sum_{j=1}^{\infty} R_{j-1} \right) \left( \sum_{j=1}^{\infty} R_{j-1}' \right) \\ &= R(1)R(1)'. \end{aligned} \quad (2.67)$$

From the definition of  $\bar{V}_s$  is it clear that the  $s$  steps ahead forecast error variance decomposition for  $x_i$ ,  $i = 1, \dots, n$ , can be performed on either  $V_s^*$  or  $\bar{V}_s$ . More importantly, the long run forecast error variance decomposition for  $x_i$  can be computed from  $\bar{V}_{\text{inf}}$ . In particular, by Theorem 2.2 we know that  $R(1)R(1)' = AA'$ .

### 2.4.3 Asymptotic Properties of Forecast Error Variance Decompositions

To my knowledge, the only papers which provide analytical expressions of the asymptotic distributions of forecast error variance decompositions are those by Lütkepohl [59] and Lütkepohl and Poskitt [61] (see also Reinsel [76], Reinsel and Lewis [77], and Yamamoto [104]). Lütkepohl [59] shows that the asymptotic variances are quite easy to calculate when the lag order in a vector autoregressive model for covariance stationary time series has a known upper bound. In case the lag order is unknown and possibly infinite, then the asymptotic variances of forecast error variance decompositions can be derived under the assumptions made by Lütkepohl and Poskitt.

The analytical expressions for the asymptotic covariance matrices which are derived in Lütkepohl [59] (and Lütkepohl and Poskitt) are, however, not based on the fact that the joint influence of all innovations account for all of the forecast error variance in any variable. Since such a fact is basically a set of linear restrictions and constrained estimators generally have smaller variances (at least in a matrix sense) than unconstrained estimators, it seems plausible that there are efficiency gains to be made from taking the restrictions into account.

Before the main results are stated and proven, some additional notation will be useful. Let  $v_{il,s}$  denote the fraction of the  $s$  steps ahead forecast error variance of  $\Delta x_i$  which is accounted for by shocks in  $\nu_l$ , where  $i, l \in \{1, \dots, n\}$ . Similarly,  $v_{il,s}^*$  is the fraction of the  $s$  steps ahead forecast error variance of  $x_i$  which is accounted for by shocks in  $\nu_l$ , whereas  $\bar{v}_{il}$  denotes the long run fraction of the forecast error variance in the levels series  $x_i$  which is accounted for by shocks in  $\nu_l$ . Accordingly,

$$v_{il,s} = \frac{\sum_{j=1}^s (e_i' R_{j-1} e_l)^2}{\sum_{j=1}^s e_i' R_{j-1} R_{j-1}' e_i},$$

for  $i, l \in \{1, \dots, n\}$  and  $s = 1, 2, \dots$ . Here,  $e_i$  is the  $i$ :th column of  $I_n$ . Furthermore, letting  $R_m^* = \sum_{j=1}^m R_{j-1}$ , we have that

$$v_{il,s}^* = \frac{\sum_{m=1}^s (e_i' R_m^* e_l)^2}{\sum_{m=1}^s e_i' R_m^* R_m^{*'} e_i},$$

for  $i, l \in \{1, \dots, n\}$  and  $s = 1, 2, \dots$ , and finally,

$$\bar{v}_{il} = \frac{(e_i' A e_{l(k)})^2}{e_i' A A' e_i},$$

for  $i \in \{1, \dots, n\}$  and  $l \in \{1, \dots, k\}$ . Here,  $e_{l(k)}$  is the  $l$ :th column of  $I_k$ .

For practical as well as theoretical reasons, it is preferable to examine the forecast error variance decompositions in terms of matrices. In particular, letting  $\odot$  denote the Hadamard product (see, e.g., Magnus and Neudecker [63, page 45]) we find that

$$v_s = \left[ \sum_{j=1}^s (R_{j-1} R_{j-1}' \odot I_n) \right]^{-1} \left[ \sum_{j=1}^s (R_{j-1} \odot R_{j-1}) \right], \quad (2.68)$$

for  $s = 1, 2, \dots$ . The  $(i, l)$ :th element of  $v_s$  is  $v_{il,s}$ . Furthermore,

$$v_s^* = \left[ \sum_{m=1}^s (R_m^* R_m^{*\prime} \odot I_n) \right]^{-1} \left[ \sum_{m=1}^s (R_m^* \odot R_m^*) \right], \quad (2.69)$$

whereas

$$\bar{v}_{\text{inf}} = [AA' \odot I_n]^{-1} [A \odot A]. \quad (2.70)$$

The  $(i, l)$ :th elements of these two matrices correspond to  $v_{il,s}^*$  and  $\bar{v}_{il}$ , respectively. It should be noted that if all elements in some row of  $A_0$  are equal to zero, this row must be deleted from  $\hat{A}$  when computing  $\hat{v}_{\text{inf}}$ . The reason is, of course, that the corresponding diagonal element of  $\hat{A}\hat{A}'$  is zero and  $[\hat{A}\hat{A}' \odot I_n]$  is consequently singular.

For any  $m \times n$  matrix  $G$ , let  $\text{diag}[\text{vec}(G)]$  denote the  $mn \times mn$  diagonal matrix whose diagonal elements are given by  $\text{vec}(G)$ . Regarding the asymptotic distributions of the estimated forecast error variance decompositions it can now be stated that:

**Theorem 2.4** *If the assumptions in Theorem 2.3 are satisfied, then*

$$T^{1/2} (\text{vec}(\hat{v}_s) - \text{vec}(v_s)) \xrightarrow{d} \mathcal{N}(0, w_s), \quad (2.71)$$

for  $s = 1, 2, \dots$ , where

$$w_s = \frac{\partial \text{vec}(v_s)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(v_s)}{\partial \beta} + \frac{\partial \text{vec}(v_s)}{\partial \omega'} V_\omega \frac{\partial \text{vec}(v_s)}{\partial \omega},$$

while

$$\begin{aligned} \partial \text{vec}(v_s) / \partial \beta' &= 2[I_n \otimes [\sum_{i=1}^s (R_{i-1} R_{i-1}' \odot I_n)]^{-1}] \sum_{j=1}^s \{ \text{diag}[\text{vec}(R_{j-1})] - \\ &\quad [v_s' \otimes I_n] \text{diag}[\text{vec}(I_n)] [R_{j-1} \otimes I_n] \} (\partial \text{vec}(R_{j-1}) / \partial \beta'), \end{aligned}$$

$$\begin{aligned} \partial \text{vec}(v_s) / \partial \omega' &= 2[I_n \otimes [\sum_{i=1}^s (R_{i-1} R_{i-1}' \odot I_n)]^{-1}] \sum_{j=1}^s \{ \text{diag}[\text{vec}(R_{j-1})] - \\ &\quad [v_s' \otimes I_n] \text{diag}[\text{vec}(I_n)] [R_{j-1} \otimes I_n] \} (\partial \text{vec}(R_{j-1}) / \partial \omega'). \end{aligned}$$

Regarding the asymptotic distribution of  $\hat{v}_s^*$ , we have that

$$T^{1/2} (\text{vec}(\hat{v}_s^*) - \text{vec}(v_s^*)) \xrightarrow{d} \mathcal{N}(0, w_s^*), \quad (2.72)$$

for  $s = 1, 2, \dots$ , where

$$w_s^* = \frac{\partial \text{vec}(v_s^*)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(v_s^*)}{\partial \beta} + \frac{\partial \text{vec}(v_s^*)}{\partial \omega'} V_\omega \frac{\partial \text{vec}(v_s^*)}{\partial \omega},$$

while

$$\begin{aligned} \partial \text{vec}(v_s^*) / \partial \beta' &= 2[I_n \otimes [\sum_{i=1}^s (R_i^* R_i^{*\prime} \odot I_n)]^{-1}] \sum_{m=1}^s \{ \text{diag}[\text{vec}(R_m^*)] - \\ &\quad [v_s^{*\prime} \otimes I_n] \text{diag}[\text{vec}(I_n)] [R_m^* \otimes I_n] \} (\partial \text{vec}(R_m^*) / \partial \beta'), \end{aligned}$$

$$\begin{aligned} \partial \text{vec}(v_s^*) / \partial \omega' &= 2[I_n \otimes [\sum_{i=1}^s (R_i^* R_i^{*\prime} \odot I_n)]^{-1}] \sum_{m=1}^s \{ \text{diag}[\text{vec}(R_m^*)] - \\ &\quad [v_s^{*\prime} \otimes I_n] \text{diag}[\text{vec}(I_n)] [R_m^* \otimes I_n] \} (\partial \text{vec}(R_m^*) / \partial \omega'). \end{aligned}$$

Finally, the asymptotic distribution for  $\hat{v}_{\text{inf}}$  is given by

$$T^{1/2} \left( \text{vec}(\hat{v}_{\text{inf}}) - \text{vec}(\bar{v}_{\text{inf}}) \right) \xrightarrow{d} \mathcal{N}(0, \bar{w}_{\text{inf}}), \quad (2.73)$$

where

$$\bar{w}_{\text{inf}} = \frac{\partial \text{vec}(\bar{v}_{\text{inf}})}{\partial \beta'} V_{\beta} \frac{\partial \text{vec}(\bar{v}_{\text{inf}})}{\partial \beta} + \frac{\partial \text{vec}(\bar{v}_{\text{inf}})}{\partial \omega'} V_{\omega} \frac{\partial \text{vec}(\bar{v}_{\text{inf}})}{\partial \omega},$$

while

$$\begin{aligned} \partial \text{vec}(\bar{v}_{\text{inf}}) / \partial \beta' &= 2[I_k \otimes [AA' \odot I_n]^{-1}] \{ \text{diag}[\text{vec}(A)] - \\ &\quad [\bar{v}'_{\text{inf}} \otimes I_n] \text{diag}[\text{vec}(I_n)] [A \otimes I_n] \} (\partial \text{vec}(A) / \partial \beta'), \\ \partial \text{vec}(\bar{v}_{\text{inf}}) / \partial \omega' &= 2[I_k \otimes [AA' \odot I_n]^{-1}] \{ \text{diag}[\text{vec}(A)] - \\ &\quad [\bar{v}'_{\text{inf}} \otimes I_n] \text{diag}[\text{vec}(I_n)] [A \otimes I_n] \} (\partial \text{vec}(A) / \partial \omega'). \end{aligned}$$

**Proof** Comparing the expressions for  $v_s$ ,  $v_s^*$  and  $\bar{v}_{\text{inf}}$  it is obvious that the asymptotic covariance matrices for estimates of these functions have a similar parametric structure. Hence, I shall limit the proof to equation (2.71).

To derive the matrices of partial derivatives of  $\text{vec}(v_s)$  with respect to  $\beta$  and  $\omega$ , note that the first differential of  $v_s$  is given by

$$\begin{aligned} dv_s &= -[\sum_{i=1}^s (R_{i-1} R'_{i-1} \odot I_n)]^{-1} [\sum_{j=1}^s d(R_{j-1} R'_{j-1} \odot I_n)] v_s + \\ &\quad [\sum_{i=1}^s (R_{i-1} R'_{i-1} \odot I_n)]^{-1} [\sum_{j=1}^s d(R_{j-1} \odot R_{j-1})] \\ &= [\sum_{i=1}^s (R_{i-1} R'_{i-1} \odot I_n)]^{-1} \sum_{j=1}^s \{ [(dR_{j-1} \odot R_{j-1}) + (R_{j-1} \odot dR_{j-1})] \\ &\quad - [(d(R_{j-1} R'_{j-1}) \odot I_n)] v_s \}. \end{aligned}$$

Taking vec's and using Magnus and Neudecker [64, Lemma 2], we obtain

$$\begin{aligned} d\text{vec}(v_s) &= [I_n \otimes [\sum_{i=1}^s (R_{i-1} R'_{i-1} \odot I_n)]^{-1}] \sum_{j=1}^s \{ 2 \text{diag}[\text{vec}(R_{j-1})] d\text{vec}(R_{j-1}) \\ &\quad - [v'_s \otimes I_n] \text{diag}[\text{vec}(I_n)] d\text{vec}(R_{j-1} R'_{j-1}) \}. \end{aligned} \quad (2.74)$$

The first differential of  $R_{j-1} R'_{j-1}$  is

$$d(R_{j-1} R'_{j-1}) = (dR_{j-1}) R'_{j-1} + R_{j-1} (dR'_{j-1}).$$

Applying the vec operator and the commutation matrix, we have that

$$\begin{aligned} d\text{vec}(R_{j-1} R'_{j-1}) &= [R_{j-1} \otimes I_n] d\text{vec}(R_{j-1}) + [I_n \otimes R_{j-1}] K_{nn} d\text{vec}(R_{j-1}) \\ &= 2N_n [R_{j-1} \otimes I_n] d\text{vec}(R_{j-1}), \end{aligned} \quad (2.75)$$

by Magnus and Neudecker [65, Lemma 4]. Substituting equation (2.75) for the differential  $d\text{vec}(R_{j-1} R'_{j-1})$  into equation (2.74), it is clear that equation (2.71) follows if

$$\text{diag}[\text{vec}(I_n)] N_n = \text{diag}[\text{vec}(I_n)] K_{nn} = \text{diag}[\text{vec}(I_n)].$$

If the second equality is valid, the first is an immediate consequence of the definition of  $N_n$ . To show the second, note that

$$\text{diag}[\text{vec}(I_n)] = \begin{bmatrix} e_1 e'_1 \otimes e'_1 \\ e_2 e'_2 \otimes e'_2 \\ \vdots \\ e_n e'_n \otimes e'_n \end{bmatrix}, \quad (2.76)$$

whereas

$$K_{nn} = \begin{bmatrix} I_n \otimes e'_1 \\ I_n \otimes e'_2 \\ \vdots \\ I_n \otimes e'_n \end{bmatrix}. \quad (2.77)$$

Since  $K_{nn}$  is symmetric (cf. Magnus and Neudecker [63, page 47]) it is immediately clear from an inspection of (2.76) and (2.77) that the result follows. Q.E.D.

In a common trends framework it may be of particular interest to study the joint influence of, e.g., the permanent versus that of the transitory innovations on the forecast error variance of the time series. Other linear functions of variance decompositions that can be relevant in empirical studies of macroeconomic time series are sums of real versus sums of nominal innovations and sums of domestic versus sums of foreign innovations. To examine such linear functions, let  $G$  be an  $n \times q$  matrix and  $\bar{G}$  be a  $k \times q$  matrix whose rows are taken from the first  $k$  rows of  $G$ . Consider the matrix functions

$$\begin{aligned} H_s &= v_s G, \\ H_s^* &= v_s^* G, \\ \bar{H}_{\text{inf}} &= \bar{v}_{\text{inf}} \bar{G}, \end{aligned} \quad (2.78)$$

for  $s = 1, 2, \dots$ . The following result can easily be proven:

**Corollary 2.2** *If the assumptions in Theorem 2.3 are satisfied, then*

$$T^{1/2} \left( \text{vec}(\hat{H}_s) - \text{vec}(H_s) \right) \xrightarrow{d} \mathcal{N}(0, [G' \otimes I_n] w_s [G \otimes I_n]), \quad (2.79)$$

$$T^{1/2} \left( \text{vec}(\hat{H}_s^*) - \text{vec}(H_s^*) \right) \xrightarrow{d} \mathcal{N}(0, [G' \otimes I_n] w_s^* [G \otimes I_n]), \quad (2.80)$$

for  $s = 1, 2, \dots$ , while

$$T^{1/2} \left( \text{vec}(\hat{H}_{\text{inf}}) - \text{vec}(\bar{H}_{\text{inf}}) \right) \xrightarrow{d} \mathcal{N}(0, [\bar{G}' \otimes I_n] \bar{w}_{\text{inf}} [\bar{G} \otimes I_n]). \quad (2.81)$$

It is claimed above that there are theoretical and practical reasons for analysing the asymptotic properties of forecast error variance decompositions in terms of matrices rather than scalars. From a practical point of view, the above matrix expressions are easier to compute than the corresponding elements obtained from an analysis of the limiting distribution for each individual variance decomposition (cf. Lütkepohl [59, Proposition 1] for such expressions). Theoretically, the above covariance matrices take into account that the rows of  $v_s$ ,  $v_s^*$ , and  $\bar{v}_{\text{inf}}$  sum to one. Accordingly, the rank of  $w_s$  and  $w_s^*$  is (less than or) equal to  $n(n-1)$ , while the rank of  $\bar{w}_{\text{inf}}$  is (less than or) equal to  $n(k-1)$ .

The reduced rank property of the covariance matrices can be stated in a very simple yet general form. Let  $\iota_m$  be the  $m \times 1$  unit vector. We then have that:

**Corollary 2.3** *If  $G = \iota_n$  and  $\bar{G} = \iota_k$ , then the asymptotic covariance matrices in Corollary 2.2 are equal to zero.*

**Proof** This follows immediately from the fact that  $v_s \iota_n = v_s^* \iota_n = \bar{v}_{\inf \iota_k} \equiv \iota_n$ .<sup>14</sup>

This result can now be used to determine the asymptotic relationship between two groups of innovations for a particular variance decomposition. In particular, let  $g$  be an  $n \times 1$  vector with known elements and  $g_\perp := \iota_n - g$ . Similarly, let  $\bar{g}$  be the  $k \times 1$  vector obtained from the first  $k$  elements of  $g$  and  $\bar{g}_\perp := \iota_k - \bar{g}$ . It now follows that

**Corollary 2.4** *The asymptotic covariance matrices in Corollary 2.2 are equal for  $G = g$  and  $G = g_\perp$  for  $s = 1, 2, \dots$  and also for  $\bar{G} = \bar{g}$  and  $\bar{G} = \bar{g}_\perp$ .*

**Proof** Below I shall show that for any  $n \times 1$  vector  $g$  with  $g_\perp := \iota_n - g$  the following equality holds true:

$$[g' \otimes I_n] w_s [g \otimes I_n] = [g'_\perp \otimes I_n] w_s [g_\perp \otimes I_n].$$

The two remaining equalities then follow from similar arguments.

The first differential of  $v_s \iota_n \equiv \iota_n$  is

$$(dv_s) \iota_n = 0.$$

Using the fact that  $\iota_n = g + g_\perp$ , we obtain

$$(dv_s) g = -(dv_s) g_\perp.$$

Taking vec's provides us with

$$[g' \otimes I_n] d\text{vec}(v_s) = -[g'_\perp \otimes I_n] d\text{vec}(v_s),$$

or in terms of partial derivatives

$$[g' \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \beta'} = -[g'_\perp \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \beta'}, \quad (2.82)$$

and

$$[g' \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \omega'} = -[g'_\perp \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \omega'}. \quad (2.83)$$

Postmultiplying both sides of (2.82) by  $V_\beta$  times the transpose of the expression in the equation we get

$$[g' \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(v_s)}{\partial \beta} [g \otimes I_n] = [g'_\perp \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \beta'} V_\beta \frac{\partial \text{vec}(v_s)}{\partial \beta} [g_\perp \otimes I_n]. \quad (2.84)$$

Similarly, for (2.83) we find that

$$[g' \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \omega'} V_\omega \frac{\partial \text{vec}(v_s)}{\partial \omega} [g \otimes I_n] = [g'_\perp \otimes I_n] \frac{\partial \text{vec}(v_s)}{\partial \omega'} V_\omega \frac{\partial \text{vec}(v_s)}{\partial \omega} [g_\perp \otimes I_n]. \quad (2.85)$$

---

<sup>14</sup>A more sophisticated proof of this corollary would be based on showing that  $[\iota'_n \otimes I_n] \text{diag}[\text{vec}(R_{j-1})] = [\iota'_n v'_s \otimes I_n] \text{diag}[\text{vec}(I_n)] [R_{j-1} \otimes I_n]$ . Such a proof is available from the author on request.

Adding the expressions in equations (2.84) and (2.85) the result follows.

Q.E.D.

Hence, letting  $G = g$  be a vector with ones in the first  $k$  elements and zeros elsewhere or vice versa, i.e.,  $G = g_{\perp}$ , will provide us with identical estimates of the asymptotic covariances matrices for  $\hat{H}_s$  and  $\hat{H}_s^*$ . That is, the standard error for an estimate of the joint influence of the permanent innovations in a variance decomposition is equal to the standard error for an estimate of the joint influence of the transitory innovations.

In essence, the property that all rows of the forecast error variance decomposition matrices sum to one is a set of restrictions. Accordingly, it can be expected that taking these restrictions into account when estimating the asymptotic covariance matrices will provide us with *efficient* estimates of these parameters in relation to estimates which ignore the  $n$  restrictions.

It should also be noted that although the above expressions are obtained from a common trends model, the results are more general than may first be obvious. In fact, equation (2.71) can be applied for calculating estimates of the asymptotic variances of variance decompositions for the known lag order vector autoregressive model which Lütkepohl [59] studies. However, the matrices  $(\partial \text{vec}(R_{j-1})/\partial \beta')$  and  $(\partial \text{vec}(R_{j-1})/\partial \omega')$  should be replaced with the expressions given in part (iv) of Proposition 1 in Lütkepohl. This solves Lütkepohl's problem (p. 120) of not taking the restrictions  $\sum_{i=1}^n v_{ii,s} = 1$  for  $i = 1, \dots, n$  into account. Unfortunately, if  $v_{ii,s} = 0$  or if  $v_{ii,s} = 1$ , the asymptotic variance of  $\hat{v}_{ii,s}$  is zero here as well as in Proposition 1 in Lütkepohl. Hence, a formal significance test of the hypothesis  $v_{ii,s} = 0$  *cannot* be conducted.<sup>15</sup>

## 2.5 Concluding Comments

In this chapter I analyse how we can estimate a common trends model with  $k$  permanent and  $r$  transitory innovations when the time series are cointegrated of order (1,1) with  $r$  cointegrating vectors. Such innovations may be of particular interest when we are interested in studying connections between growth and business cycle fluctuations in macroeconomic time series. Theorem 2.1 establishes that all reduced form parameters can be calculated directly from the (estimated) parameters of a restricted vector autoregression. The restrictions this model satisfies are the cointegration constraints which may, e.g., describe the steady state of a stochastic growth model. Furthermore, a simple description of the solution to the unrestricted and restricted vector autoregressions as well as the vector error correction representation in terms of estimable parameters is provided. The solution is expressed as a vector moving average representation which describes a reduced form of the propagation mechanism.

Second, an identification matrix for permanent and transitory innovations has been derived. Since this matrix is generally not triangular and is based on certain assumptions regarding the nature of permanent and transitory innovations, it cannot be computed by a Choleski, an eigen, or a method of moments decomposition of  $\Sigma$ . However, it is shown that its parameters are solely determined by parameters which have already been calculated and is therefore easily obtained in practise.

---

<sup>15</sup>Obviously, this is also true for  $v_{ii,s}^*$  and  $\bar{v}_{ii}$ .

Third, I have analysed the asymptotic properties of impulse response functions and forecast error variance decompositions within a common trends model when an upper bound for the lag order and the cointegration matrix  $\alpha$  are known. Such functions are calculated directly from the vector moving average representation and the identification matrix. Based on the results in Theorems 2.3 and 2.4, we find that the analytical expressions of the asymptotic covariances are somewhat more complex than in the model studied by Lütkepohl and Reimers [62]. The reason for the added complexity is that I have analysed identification of permanent and transitory innovations, whereas Lütkepohl and Reimers study the case when the covariance matrix  $\Sigma$  is orthogonalized via a Choleski decomposition. From a practical point of view, however, the added complexity is not severe. For example, to compute estimates of the asymptotic covariance matrices of impulse response functions is about as time consuming as for ordinary vector autoregressions. Finally, Corollary 2.2 provides us with asymptotic distributions of the estimated forecast error variances which are accounted for by linear functions of the innovations. Based on these distributions it is, e.g., possible to analyse how important innovations to growth are, at a business cycle horizon, relative to transitory shocks for the time series of interest. Also, it highlights the fact that the asymptotic covariance matrices of the estimated forecast error variance decompositions are conditioned on the property that the sum of a decomposition for any time series is equal to one.

# Chapter 3

## Common Trends and Macroeconomic Fluctuations

### 3.1 Introduction

The purpose of this chapter is to illustrate the theory in Chapter 2. I shall consider data on six macroeconomic time series: terms of trade, the money stock, the price level, real output, and real consumption. The choice of variables is based on a small open economy real business cycle model due to Lundvik [57] and a steady state solution to this model will be used to derive cointegration vectors.

The particular data sets I shall examine have been collected annually; for Finland (1866–1985) and for Sweden (1871–1986). The discussion on the estimation of common trends models for these samples will be related to the following questions:

- (i) the relative importance of permanent and transitory shocks,
- (ii) the relative importance of (permanent) domestic and foreign shocks,
- (iii) the relative importance of (permanent) real and nominal shocks, and
- (iv) whether the sources of fluctuations are similar in the two countries.

It should be emphasized that I view *all* conclusions reached on these questions here as illustrations. Whether they may also be considered “empirical regularities” will not be addressed in this thesis.

The remainder of the chapter is organized as follows. Section 3.2 contains a discussion of some general issues involved when analysing macroeconomic fluctuations. Theoretical cointegration constraints suggested by an open economy growth model are discussed in section 3.3. Finally, the illustrations for the Finnish and the Swedish data are presented in section 3.4 and some concluding remarks in section 3.5.

### 3.2 Analysis of Macroeconomic Fluctuations

The two illustrations in section 3.4 will be based on three ideas which may capture some important characteristics of macroeconomic fluctuations. First, such fluctuations can be

well modelled as being driven by stochastic shocks to the economy. The shocks themselves do not follow any regular, cyclical processes; if they did it would be possible to predict them, in which case they would cease to act as pure shocks to the economy. The cyclicity, instead, derives from various propagation mechanisms, through which the disturbances are transmitted. Second, the secular trends may also be represented by a stochastic process, just like the short run cycles. Third, the number of independent trends is rather low compared to the number of relevant macroeconomic variables.

The impulse-propagation mechanism idea is well captured by a vector moving average model of the data generating process:

$$x_t = \bar{x}_t + H(L)\varepsilon_t, \quad (3.1)$$

where  $x_t$  is an  $n \times 1$  vector time series and  $\bar{x}_t$  contains the deterministic components of  $x_t$ ,  $L$  is the lag operator (i.e.,  $L^j x_t = x_{t-j}$ ),  $H(\lambda)$  is an  $n \times n$  matrix polynomial, and the  $n \times 1$  vector  $\varepsilon_t$  is purely nondeterministic and serially uncorrelated with mean zero and positive definite covariance matrix  $\Sigma$ . The matrix function  $H(\lambda)$  thus represents the mechanism through which the impulses  $\varepsilon_t$  are propagated.

The vector moving average model (3.1) is a natural starting point for analyses of the contributions of different types of disturbances to movements in  $x_t$ . In the terminology of vector autoregressions, an *impulse response function* gives the response of an element of  $x_{t+s}$ ,  $s = 0, 1, \dots$ , to an unpredicted movement in some component of  $\varepsilon_t$ ; the impulse response functions are thus given directly by  $H(\lambda)$ . A *variance decomposition* of the  $s$  steps ahead forecast error variance gives the proportion of the total forecast error variance of one component of  $x_{t+s}$  accounted for by various components of  $\varepsilon$ . It should be noted, however, that the  $\varepsilon_{i,t}$ 's typically cannot be interpreted as "structural shocks" since, among other things,  $\Sigma$  will in general not be a diagonal matrix. That is, the elements of  $\varepsilon_t$  will not have the property of being mutually uncorrelated, conventionally associated with structural shocks. The question of how impulse responses and variance decompositions associated with structural shocks can be derived and analysed is addressed in Chapter 2.

To formalize the second idea, that  $x_t$  consists of both cyclical and secular components, we may reformulate the data generating process as

$$x_t = x_0 + A\tau_t + \tilde{C}(L)\varepsilon_t, \quad (3.2)$$

where  $x_0$  is a vector of constants containing the initial values of  $x_t$ , and  $A\tau_t$  and  $\tilde{C}(L)\varepsilon_t$  capture the (wide sense) nonstationary and stationary components of  $x_t$ , respectively. It is assumed that  $\varepsilon_s = 0$  for  $s \leq 0$  and, in order for  $\tilde{C}(L)\varepsilon_t$  to be stationary, that the polynomial  $\tilde{C}(\lambda)$  is finite for all  $|\lambda| \leq 1$ . If the trends are linearly deterministic, then  $\tau_t = \mu t$ , i.e.,  $\tau_t - \tau_{t-1} = \mu$ , where  $\mu$  is a vector of constants. The idea of linear stochastic trends, on the other hand, can be operationalized by modelling  $\tau_t$  as a vector of random walks with drift:

$$\tau_t = \mu + \tau_{t-1} + \varphi_t, \quad (3.3)$$

where  $\varphi_t$  is a vector of structural shocks with permanent effects. It is, of course, not necessary to model stochastic trends as random walks with drift. However, the random walk model is simple *and* if, for example, the elements of  $\mu$  are large in relation to the variances of the elements of  $\varphi$ , then the trends will display a pattern of slowly time varying slopes. This is a property that may reflect the development in macroeconomic time series

more accurately than linear deterministic trends with constant slope parameters (cf. Stock and Watson [96]).

In practical work, it is virtually impossible to statistically distinguish a random walk with drift from a linear deterministic trend<sup>1</sup> (cf. Christiano and Eichenbaum [16], Haldrup [36], and Haldrup and Hylleberg [37]). Whether one prefers to model macroeconomic time series as being governed by random walks with drift or linear deterministic trends is therefore very much a question of taste. Most univariate empirical studies (see, e.g., Perron [72]) are consistent with the stochastic trend hypothesis, however, which indicates that an investigation of the relative importance of permanent and transitory shocks should be worthwhile.

The idea that there are fewer trends than variables can be modelled by letting the dimension of  $\tau_t$  be smaller than that of  $x_t$ . For instance, technology and preferences presumably suggest that, e.g., output and investments, or output and consumption, do not follow independent growth paths. It is thus assumed that  $\tau_t$  is a  $k \times 1$  vector, with  $k < n$ . As shown by Stock and Watson [95] (see also the example in Chapter 1), this implies that there exists  $n - k$  linearly independent linear functions of the elements of  $x_t$  which are stationary, even if the individual elements themselves are all nonstationary.

With varying success, attempts have been made to integrate nonstructural models like (3.1)–(3.3) with theoretical structures. King, Plosser, Stock and Watson [53] discuss a simple, stochastic growth model of a closed economy where consumption, investments, and output have a common trend. This trend arises from the technology factor of an aggregate Cobb–Douglas production function, where (the natural logarithm of) technology is modelled as a random walk with drift. They show that the model may be written in the form of (3.2)–(3.3). When they extend the vector of variables to include the general price level and the money stock, they assume, essentially on *a priori* grounds rather than by means of theoretical arguments, that there are two underlying stochastic trends, which they interpret as a real and a nominal trend. With two common trends and five trending variables (the logarithms of consumption, investments, output, money, and the price level) it should be possible to find three stationary linear combinations of the (log) levels. King, Plosser, Stock and Watson thus assume that the logarithms of the ratios between consumption and output and investments and output are stationary (which is consistent with their growth model) and that the logarithm of the velocity of money is also stationary.

I am not aware of any attempt to extend the analysis of King, Plosser, Stock and Watson to an open economy framework.<sup>2</sup> However, similar models have been discussed in slightly different contexts. For my purpose, the two most interesting examples are given by the open economy, deterministic growth models constructed by Lundvik [57], and Persson and Svensson [73].<sup>3</sup> In both these models the effects of variations in terms of trade on consumption, investments, and output are considered.

---

<sup>1</sup>That is, using classical test procedures. Bayesian frameworks for testing for unit roots have been presented by, e.g., DeJong and Whiteman [20], and Sims and Uhlig [93].

<sup>2</sup>In an empirical analysis of the British balance of trade, Ahmed [1] treats the main forcing variable, real government (military) spending, as a random walk with drift. The theoretical framework is not stochastic, however.

<sup>3</sup>In spite of the fact that growth is deterministic in these models, they are not qualitatively different from the model in King, Plosser, Stock and Watson [53].

Lundvik's model has the property that from its steady state solution two (log) linear combinations between terms of trade, consumption, investments, and output can be derived. If the four variables are driven by stochastic trends and if stationary (transient) fluctuations around the steady state values are allowed for, the two relationships can be taken to indicate the existence of two common trends. I will discuss the implications of Lundvik's model in some detail in section 3.3 below.

### 3.3 Cointegration Vectors and Common Trends in a Small Open Economy Growth Model

Models like (3.1) or (3.2)–(3.3) are not easily estimated since they are formulated in terms of unobserved components. The approach suggested by Sims [90] is to estimate a vector autoregression, which can be inverted with standard techniques to yield a moving average model like (3.1) provided that  $\{x_t\}_{t=1}^{\infty}$  is jointly covariance stationary. On the other hand, if we wish to model data as being driven by a few common stochastic trends in the sense discussed in section 3.2, standard inversion algorithms of unrestricted vector autoregressions (UVAR) will no longer be valid. On the basis of the analysis by Engle and Granger [26] and Stock and Watson [95], King, Plosser, Stock and Watson [53] suggest that we can estimate and invert a vector error correction model. Alternatively, a restricted vector autoregression (RVAR) can be estimated and inverted along the lines suggested in Chapter 2. I consider the latter procedure preferable from a practical point of view as it, e.g., yields a simpler inversion algorithm.

As discussed in section 2.2.2, the implementation of a restricted vector autoregression requires an estimate of the cointegrating matrix. Given  $\alpha$ , we may then proceed and calculate the common trends model (3.2), which requires some further information about the structure of  $A_0$  and  $\pi$  (where  $A = A_0\pi$ ). In this section I shall present the assumptions about  $\alpha$  and  $A_0$  that will be used in the empirical analyses in section 3.4.

In the absence of any *prior* information the methods in Johansen [44,45,48], and Johansen and Juselius [49] still make it possible (i) to test for the rank of  $\alpha$ , i.e., the number of cointegrating vectors, and (ii) for a given rank also a numerical structure of these vectors. However, economic theory should be able to provide information on both these issues. Use of such information will greatly facilitate the interpretation of the common trends model and, moreover, probably increase the efficiency of the resulting estimates. Since Lundvik's [57] real business cycle model of a small open economy is of relevance for this chapter I will consider certain implications of his model in some detail.

Below I shall derive two log-linear relationships among real output, real consumption, real investments, and terms of trade from the certainty equivalent steady state solution to Lundvik's model. Since the steady state solution variables are detrended transformations of the original variables and, moreover, the steady state solution has been calculated in the absence of any stationary stochastic disturbances, the derived log-linear relationships do in general not yield sufficient information about possible cointegrating vectors. To be able to interpret the relationships as corresponding directly to cointegrating vectors I shall make two additional assumptions which, however, turn out not to be very restrictive. First, it is assumed that the variables of interest can be multiplicatively decomposed into a nonstationary stochastic trend term and a stationary component (and, hence, that the

logarithms of the same variables can be additively decomposed). This assumption is made in virtually all studies employing the cointegration idea. Second, I suppose that it is permissible to view the steady state solution as pertaining to a situation in which all stationary disturbances in the system are equal to their expected, zero, values.

To distinguish the steady state solution variables from the corresponding observed macro variables I will denote the former by small letters and the latter by capitals. Output,  $y$ , is produced by means of a Cobb–Douglas technology. There are three factors of production: labor and two types of goods, the capital stocks of which are denoted by  $k_h$  and  $k_f$ . The labor input being normalized to unity, the production function can be written as

$$y = ak_h^{\theta_h} k_f^{\theta_f}, \quad (3.4)$$

where  $a$  is a constant,  $\theta_h, \theta_f > 0$  and  $\theta_h + \theta_f < 1$ . The index  $h$  denotes a good that is produced in the home country and abroad, while the index  $f$  indicates that the good is only produced abroad, in the foreign country. Letting the price of the home country good act as *numeraire*, maximization of (3.4) subject to a cost constraint yields the following first order condition for the allocation of the two capital goods

$$k_f = t(\theta_f/\theta_h)k_h, \quad (3.5)$$

where  $t$  is terms of trade.<sup>4</sup> According to the steady state solution,  $k_h$  can be expressed as

$$k_h = v(at^{\theta_f})^{1/(1-\theta_h-\theta_f)}, \quad (3.6)$$

where  $v$  is a constant. Total investments,  $i$ , on the other hand, is given by

$$i = i_h + i_f = (g + d)k_h + (g + d)\frac{k_f}{t}, \quad (3.7)$$

where  $g$  is the rate of growth in output and  $d$  the rate of depreciation. Inserting (3.5) and (3.6) into (3.4) and (3.7) we obtain after some manipulations

$$\ln y - \ln i = c^*, \quad (3.8)$$

where  $c^*$  is a constant involving  $v$ ,  $\theta_h$ ,  $\theta_f$ ,  $g$ , and  $d$ . From this expression it can be seen that even if the corresponding observed macro variables,  $Y$  and  $I$ , are driven by stochastic trends,  $\ln Y - \ln I$  will be stationary.

Let us now turn to a linear relationship between consumption, output, investments and terms of trade. The steady state budget constraint can be written

$$c = y - [1 + h(g + d)]i + (r - g)p^c b, \quad (3.9)$$

where  $c$  is aggregate (detrended) consumption,  $h$  is a parameter determining the installation cost associated with new investments,  $r$  is the real rate of interest,  $p^c$  is the price of aggregate consumption, and  $p^c b$  is bond holdings, nominated in terms of aggregate consumption. Lundvik employs a CES function to aggregate consumption of the two goods,

---

<sup>4</sup>Regarding terms of trade, my notation differs from Lundvik's. Instead of defining terms of trade in the conventional way Lundvik uses its inverse, i.e., the ratio between the import price index and the export price index, which he denotes by  $p$ . Thus, the  $t$  that I use is the inverse of Lundvik's  $p$ .

$c_h$  and  $c_f$ , to the composite good,  $c$ . For computational simplicity let us instead assume that the aggregator function is Cobb–Douglas. In this case  $p^c$  is given by

$$p^c = \beta^{-\beta}(1 - \beta)^{\beta-1}t^{\beta-1}, \quad (3.10)$$

where  $\beta$  is the share of  $c_h$  in total consumption.

Assuming that in steady state the stock of bonds grows at the same rate as output we can substitute  $\kappa y$  for  $b$  in (3.9), where  $\kappa$  is a constant of proportionality. Together with (3.10) this makes it possible to rewrite (3.9) according to

$$\begin{aligned} c &= y - [1 + h(g + d)]i + [\beta^{-\beta}(1 - \beta)^{\beta-1}\kappa]t^{\beta-1}y \\ &\equiv y - m_1i + m_2t^{\beta-1}y. \end{aligned} \quad (3.11)$$

In order to express (3.11) in log–linear form, let us approximate it by the corresponding constant elasticity relation.<sup>5</sup> Doing so, we obtain

$$\begin{aligned} \ln c &= E[y/c] \ln y - m_1 E[i/c] \ln i + m_2 E[t^{\beta-1}y/c] \ln t^{\beta-1}y \\ &= -\alpha_{31} \ln t - (\alpha_{34} - \alpha_{32}) \ln y - \alpha_{35} \ln i, \end{aligned} \quad (3.12)$$

where the somewhat odd notation will be motivated below.

The two cointegrating vectors which have been derived here relate only to real aggregates. When extending the model to include nominal variables in the form of the money stock ( $M_t$ ) and the general price level ( $P_t$ ) I will follow King, Plosser, Stock and Watson and impose the constraint that the logarithm of the velocity of money, i.e.,  $(\ln Y_t + \ln P_t - \ln M_t)$ , is stationary.

Taken together we thus have three cointegrating vectors relating to six variables;  $\ln T_t$ ,  $\ln P_t$ ,  $\ln M_t$ ,  $\ln Y_t$ ,  $\ln I_t$ , and  $\ln C_t$ . Given this ordering of the variables, the  $\alpha$  matrix can be written

$$\alpha' = \begin{bmatrix} 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ \alpha_{31} & 0 & 0 & (\alpha_{34} - \alpha_{32}) & \alpha_{35} & 1 \end{bmatrix}, \quad (3.13)$$

where I have ordered the cointegrating vectors such that the first and second one derived constitute the second and third rows of  $\alpha'$ . The ordering of the *rows* of  $\alpha'$  is immaterial and is just a matter of convenience. However, to ensure that the matrix  $M$  (cf. section 2.2.1) is nonsingular, the *columns* of  $\alpha'$  have been ordered such that the final three columns make up a nonsingular matrix.<sup>6</sup>

When it comes to empirical implementation of the  $\alpha$  matrix a problem arises with respect to its last column, since its coefficients have to be estimated in one way or another.

---

<sup>5</sup>To this end, we can make use of two simple facts. First, that the elasticity of a variable  $x$  with respect to another variable  $y$  is given by  $e_{xy} := (\partial x / \partial y)(y/x) = \partial \ln x / \partial \ln y$ . Secondly, that, up to a first order approximation, the expected value of a product of two stochastic variables,  $a$  and  $b$  say, is equal to the product of their respective expectations, i.e.,  $E[ab] \approx E[a]E[b]$ . Thus, if  $a = \partial x / \partial y$  and  $b = x/y$ , we can estimate  $E[e_{xy}]$  by estimating  $E[a]$  and  $E[b]$  separately and multiply these two estimates. In terms of (3.11)  $E[a]$  corresponds to 1,  $m_1$ , and  $m_2$ , respectively. The expected values of  $b$  correspond to the expected values of the ratios  $y/c$ ,  $i/c$ , and  $t^{\beta-1}y/c$ , which can be estimated by the mean values of these ratios.

<sup>6</sup>This is not necessary if the  $S_3$  matrix is chosen such that its linearly independent rows are orthogonal to the columns of  $\alpha$ , i.e.,  $S_3 = \alpha'_1$ , where  $\alpha'_1 \alpha = 0$  and  $\alpha'_1 \alpha_1 = I_3$  (cf. Chapter 2).

In empirical applications of his model, Lundvik [57] has assigned numerical values to the parameters in (3.11). However, to determine the coefficients in (3.12) we would, in addition, have to estimate the mean values of the ratios  $(y/c)$ ,  $(i/c)$ , and  $(t^{\beta-1}y/c)$ . Since these ratios concern the steady state solution variables,  $y$ ,  $c$ ,  $i$  and  $t$  whereas I only have data on the actual realisations of these macro variables, i.e., the time series on  $Y$ ,  $C$ ,  $I$  and  $T$ , I cannot estimate the ratios directly. Using the theory in Johansen and Juselius [49] it is possible, however, to estimate the space spanned by the last column of  $\alpha$  conditional on the other two cointegrating vectors being fully specified. For computational reasons, it is then easier to consider the following equivalent specification of the theoretical cointegrating matrix

$$\alpha' = \begin{bmatrix} 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ \alpha_{31} & \alpha_{32} & -\alpha_{32} & \alpha_{34} & \alpha_{35} & 1 \end{bmatrix}. \quad (3.14)$$

This matrix is obtained by premultiplying the cointegrating matrix in (3.13) by the nonsingular matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha_{32} & 0 & 1 \end{bmatrix}.$$

The matrix  $\alpha$  is unique up to a nonsingular linear transformation, i.e., if  $z_t$  is jointly wide sense stationary then  $a^*z_t$  is also jointly stationary for some  $3 \times 3$  matrix  $a^*$ . The requirement of nonsingularity is needed to preserve the cointegrating properties of the system. The particular transformation here, which is chosen for computational reasons, is the motivation for the somewhat odd notation in (3.12) above. More importantly, it highlights that the coefficients on the money stock and the price level in all cointegrating vectors sum to zero, i.e., that only the real money stock matters in steady state.

Let us now turn to the question about the structure of  $A_0$ . As the number of common trends is equal to the number of variables (six) minus the number of cointegrating vectors (three), we have that there are three common trends. I shall denote these trends the foreign trend ( $\tau_{f,t}$ ), the real domestic trend ( $\tau_{r,t}$ ), and the nominal domestic trend ( $\tau_{n,t}$ ), respectively. In order to interpret the trends in this way two assumptions may be considered. The first is that shocks to the domestic trends have *no* long run effect on terms of trade, i.e., the small open economy hypothesis is imposed as a long run restriction. The second assumption states that money is neutral in the long run, i.e., shocks to the nominal trend have *no* long run effects on the real variables. The following  $A_0$  matrix is consistent with these assumptions and with the restriction that  $\alpha'A_0 = 0$ :

$$A_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ a_{0,61} & a_{0,62} & 0 \end{bmatrix}, \quad (3.15)$$

where  $a_{0,61} = \alpha_{32} - (\alpha_{31} + \alpha_{34} + \alpha_{35})$ , and  $a_{0,62} = \alpha_{32} - (\alpha_{34} + \alpha_{35})$ . The zeros in the first row of  $A_0$  reflect the small open economy assumption; the foreign trend is thus the

first stochastic trend. The zeros in the third column similarly reflect the assumption of neutrality; the nominal trend is the third stochastic trend. The zeros in the second row, on the other hand, are just a convenient normalization and does *not* influence the determination of the elements of  $A$ . The structure of  $A$  in (3.2) may now be found by postmultiplication of  $A_0$  by the  $3 \times 3$  lower triangular matrix  $\pi$ . Letting  $\pi_{ij}$  denote the  $(i, j)$ :th element of  $\pi$ , we have that

$$A = \begin{bmatrix} \pi_{11} & 0 & 0 \\ \pi_{31} & \pi_{32} & \pi_{33} \\ \pi_{11} + \pi_{21} + \pi_{31} & \pi_{22} + \pi_{32} & \pi_{33} \\ \pi_{11} + \pi_{21} & \pi_{22} & 0 \\ \pi_{11} + \pi_{21} & \pi_{22} & 0 \\ a_{0,61}\pi_{11} + a_{0,62}\pi_{21} & a_{0,62}\pi_{22} & 0 \end{bmatrix}. \quad (3.16)$$

Since  $\pi$  is lower triangular and has rank 3 it is clear that  $\pi_{ii} \neq 0$  for all  $i \in \{1, 2, 3\}$ . However,  $\pi_{ij}$  is not constrained for  $i \neq j$ . It thus follows that, e.g.,  $\tau_{f,t}$  need not have a significant long run impact on any variables but the terms of trade.<sup>7</sup>

In closing, it seems appropriate to refer back to the first paragraph of this section and consider what properties of the  $\alpha$  and  $A_0$  matrices that economic theory has helped to establish. We have found that: (i)  $\text{rank}[\alpha] = 3$ , (ii) the exact numerical structure of two of the columns of  $\alpha$ , (iii) the space spanned by the remaining cointegrating vector, and (iv) the structure of  $A_0$  and hence of the loading matrix  $A$  in the common trends model.

### 3.4 Common Trends and Macroeconomic Fluctuations in Finland and Sweden

In this section I will estimate and analyse common trends models using annual data from Finland (1866–1985) and Sweden (1871–1986). The estimation strategy follows the procedures discussed in Chapter 2. It should be emphasized that the material presented below is an illustration of how one can conduct empirical analysis of macroeconomic data by applying the common trends idea. It is not my intention here to argue that the results should or could be viewed as empirical regularities of the Finnish and the Swedish economies.

#### 3.4.1 The Data

For both countries, the vector of variables is given by

$$x_t = \left[ \ln T_t \quad \ln P_t \quad \ln M_t \quad \ln Y_t \quad \ln I_t \quad \ln C_t \right]',$$

where  $T_t$  denotes terms of trade (export price index/import price index),  $P_t$  the gross domestic product (GDP) price deflator,  $M_t$  the nominal money stock per capita ( $M_2$ ),  $Y_t$

---

<sup>7</sup>Furthermore, the matrix  $A$  in (3.16) satisfies 12 restrictions and the common trends are thus exactly identified. These 12 restrictions are given by the 5 zeros and 7 of the 9 restrictions from the matrix  $\alpha'A = 0$ . The remaining 2 restrictions from this  $3 \times 3$  matrix are already implied by the zeros. Since no more than 6 elements of  $\pi$  can be uniquely determined it follows that when  $A$  (and thus  $A_0$ ) has 18 elements, we must impose 12 restrictions to exactly identify this matrix.

real GDP per capita,  $I_t$  real gross domestic investments per capita, and  $C_t$  real (private and public) consumption per capita.<sup>8</sup>

The following data sources have been used. Terms of trade for Sweden have been calculated from export and import price indices given in Ohlsson [70] for the period 1871–1966. The implicit price indices in the official national accounts have been used for the period 1967–1986. For Finland, export and import prices have been taken from Hjerpe [41]. The sources for  $I_t$ ,  $C_t$ ,  $Y_t$ , and  $P_t$  are for Sweden Krantz and Nilsson [54] (1871–1949) and the official national accounts (1950–1986), and for Finland Hjerpe [41]. I have used series in current prices from the sources, and calculated the real series simply by deflating with  $P_t$ . It should be noted that

- In order to get Sweden’s GDP at market prices 1871–1949, I have used the national income identity, and aggregate series on public and private consumption, investments, etc., in Krantz and Nilsson [54]. That is, I have *not* used their series on “domestic product”, which is given in terms of factor prices.
- In the Krantz and Nilsson [54] data set, changes in inventories are included partly in consumption and partly in investments. In the official national accounts, however, changes in inventories are treated separately (and have not been included in my series on  $I_t$ , 1950–1986).
- The GDP price deflator for Sweden is the ratio between domestic product in current and fixed factor prices 1871–1949, and between GDP in current and fixed market prices 1950–1986.

In these respects the choice of data is in line with the suggestions made by Englund, Persson and Svensson [29].

The sources for the series on population are for Sweden “Statistisk Årsbok” (various issues) and for Finland, Hjerpe [41]. The series on the Swedish money stock is the same as that used by Bergman and Jonung [7] and the series on the Finnish money stock has been obtained from Tarmo Haavisto, University of Lund.

### 3.4.2 Illustrations

Before any tests for cointegration can be performed, we have to decide on a suitable lag order of the unrestricted vector autoregression in (2.2) of the Finnish and Swedish data sets. In Table 3.1 the results are presented.

In Paulsen [71] it is shown that the log criterion (SIC) of Schwarz [84] and the iterated log criterion (ILC) of Hannan and Quinn [39] are consistent measures for determining the true lag order in the presence of unit roots (stochastic trends). The Akaike information criterion (AIC) (cf. Akaike [2]), however, is not consistent and tends to overestimate the true lag order. These findings are, of course, consistent with what has previously been established for vector autoregressions of covariance stationary time series (see, e.g., Geweke and Meese [33, Theorem 5], Quinn [74, Corollary], and Shibata [88]). Furthermore, the Monte Carlo study in Jacobson [43] suggests that the iterated log criterion tends to pick the true lag order with greater accuracy than the other two for cointegrated systems (for a

---

<sup>8</sup>The availability of money stock data is the reason I look at M2 rather than, e.g., M1.

Table 3.1: Lag order selection for UVAR( $p$ ) models on the Swedish and the Finnish data.

Information Criteria:

	lags	1	2	3	4	5	6	7	8
Sweden	AIC	-37.422	-38.173	-38.839*	-38.418	-37.854	-38.647	-38.325	-37.908
	SIC	-34.451*	-33.191	-30.803	-28.285	-25.580	-23.187	-20.633	-17.937
	ILC	-36.447	-37.290*	-36.870	-36.429	-35.829	-35.569	-35.178	-34.675
Finland	AIC	-33.608	-33.869	-34.783	-34.548	-34.139	-35.037*	-34.685	-34.382
	SIC	-30.717*	-29.047	-26.991	-24.746	-22.285	-20.089	-17.600	-15.115
	ILC	-32.662	-32.962*	-32.901	-32.678	-32.265	-32.143	-31.757	-31.403

Lag Order Tests:

Sweden	1 vs. 2	1 vs. 3	2 vs.3	Finland	1 vs. 2	1 vs. 3	2 vs. 3
LR	198.514	263.981	64.242	LR	172.306	244.972	74.610
(p-value)	(.000)	(.000)	(.003)	(p-value)	(.000)	(.000)	(.000)
W	293.221	402.497	72.451	W	235.568	338.875	82.527
(p-value)	(.000)	(.000)	(.000)	(p-value)	(.000)	(.000)	(.000)

Note: The \* indicates the minimum for a maximum lag order of 8. LR denotes the likelihood ratio test and W the Wald test. For a null hypothesis of  $p_0$  lags in a vector autoregression with  $p_1 > p_0$  lags, the limiting distribution of these tests is  $\chi^2$  with  $36(p_1 - p_0)$  degrees of freedom (see Sims, Stock and Watson [92]).

Monte Carlo study on the small sample properties of various lag order selection statistics in covariance stationary vector autoregressive systems, see Nickelsburg [68]). Based on this criterion, a lag order of 2 should be selected in the unrestricted vector autoregressions for both the Swedish and the Finnish data. On the other hand, the lag order tests (Likelihood Ratio and Wald) strongly suggest that there is, e.g., serial correlation left in the estimated residuals of a UVAR(2) model. I have nevertheless decided to let the lag order equal 2 for both countries.<sup>9</sup>

The next step is to test for the presence of stochastic trends conditional on  $p$ . In the upper panel of Table 3.2 we find the results from the Johansen [44,45,48] likelihood ratio tests for cointegration. The null hypothesis is that there are  $r$  or fewer cointegrating vectors. The alternative hypotheses are  $r + 1$  and at least  $r + 1$  cointegrating vectors for the  $LR_{\max}$  and  $LR_{tr}$  tests, respectively. For both Sweden and Finland, there is evidence of two cointegrating vectors according to the  $LR_{tr}$  test. The tests, however, also suggest that there may be as many as four cointegrating vectors for the Finnish data, according to the  $LR_{tr}$  test, while there may be no more than one in the case of Sweden, according to the  $LR_{\max}$  test.

---

<sup>9</sup>When a maximum lag order greater than 10 is examined, the information criteria tend to favor the maximum and the lag order tests always reject the null hypothesis. The primary reason why I consider the low lag order here is based on computational convenience. For example, to compute standard errors for impulse response functions and forecast error variance decompositions the matrices of partial derivatives  $(\partial \text{vec}(R_j) / \partial \beta')$  for  $j = 0, 1, \dots$  must be estimated. Every such matrix is of dimension  $36 \times 36p$ , i.e., has  $1296p$  elements. Clearly, the smaller  $p$  is the faster the computation is.

Table 3.2: Cointegration tests on the Swedish and the Finnish data for a null hypothesis of  $r$  or fewer cointegrating vectors when  $p = 2$ .

	$r$	$\hat{\lambda}_{r+1}$	$LR_{\max}$	p-value	$LR_{tr}$	p-value
Sweden	0	.506	80.334	< .01	147.467	< .01
	1	.213	27.272	.20	67.133	.05
	2	.181	22.734	.17	39.861	.20
	3	.102	12.331	.50	17.127	> .50
	4	.039	4.586	> .50	4.796	> .50
	5	.002	.210	> .50	.210	> .50
Finland	0	.401	60.413	< .01	170.856	< .01
	1	.365	53.691	< .01	110.444	< .01
	2	.193	25.288	.08	56.753	< .01
	3	.141	17.955	.14	31.465	.04
	4	.087	10.784	.17	13.510	.11
	5	.022	2.725	.11	2.725	.11

**Estimated Cointegrating Vectors:**

$$\hat{\alpha}^{(S)'} = \begin{bmatrix} -.138 & -.072 & .047 & 1.0 & -.154 & -.851 \\ -.233 & -.361 & .225 & 1.0 & .509 & -1.590 \\ .596 & .260 & -.381 & -.121 & -.238 & 1.0 \end{bmatrix}$$

$$\hat{\alpha}^{(F)'} = \begin{bmatrix} .099 & .003 & .056 & 1.0 & -.395 & -.823 \\ .827 & .308 & -.404 & -.126 & -.280 & 1.0 \\ -.085 & -.131 & .072 & 1.0 & .089 & -1.118 \end{bmatrix}$$

**Testing the Hypothesis:**

$$\alpha^{(i)'} = \begin{bmatrix} .0 & 1.0 & -1.0 & 1.0 & .0 & .0 \\ .0 & .0 & .0 & 1.0 & -1.0 & .0 \\ \alpha_{31}^{(i)} & \alpha_{32}^{(i)} & \alpha_{33}^{(i)} & \alpha_{34}^{(i)} & \alpha_{35}^{(i)} & 1.0 \end{bmatrix} \quad i = S, F.$$

	Sweden	Finland
LR(6)	76.259	85.163
p-value	(.000)	(.000)
$\hat{\alpha}_{31}^{(i)}$	.175	.139
$\hat{\alpha}_{32}^{(i)}$	.196	.200
$\hat{\alpha}_{33}^{(i)}$	-.196	-.200
$\hat{\alpha}_{34}^{(i)}$	-.385	-.407
$\hat{\alpha}_{35}^{(i)}$	-.392	-.363

**Note:** The p-values for the  $LR_{\max}$  (the alternative hypothesis is  $r + 1$  cointegrating vectors) and the  $LR_{tr}$  (the alternative being at least  $r + 1$  cointegrating vectors) are based on the critical values reported by Johansen [45] in Table T.I. The p-values for testing linear restrictions on  $\alpha$  are taken from the  $\chi^2$  distribution with 6 degrees of freedom (see Johansen and Juselius [49, Theorem 2]). The number of degrees of freedom are equal to the number of stochastic trends ( $n - r = 3$ ) times the number of known cointegrating vectors ( $s = 2$ ). In addition, we can test  $\alpha_{32}^{(i)} = -\alpha_{33}^{(i)}$  along with the known cointegrating vectors. Such a hypothesis has 7 degrees of freedom in the present case.

Table 3.3: Estimated common trends models for Sweden and Finland based on the theoretical cointegrating vectors.

**Sweden:**

$$\begin{bmatrix} \ln T_t \\ \ln P_t \\ \ln M_t \\ \ln Y_t \\ \ln I_t \\ \ln C_t \end{bmatrix} = x_0 + \begin{bmatrix} .0639 & .0 & .0 \\ (.0042) & (-) & (-) \\ -.0660 & -.0171 & .1526 \\ (.0151) & (.0144) & (.0102) \\ -.0575 & .0142 & .1526 \\ (.0149) & (.0144) & (.0102) \\ .0084 & .0314 & .0 \\ (.0030) & (.0021) & (-) \\ .0084 & .0314 & .0 \\ (.0030) & (.0021) & (-) \\ -.0030 & .0305 & .0 \\ (.0029) & (.0020) & (-) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{f,t} \\ \hat{\tau}_{r,t} \\ \hat{\tau}_{n,t} \end{bmatrix} + \hat{u}_t.$$

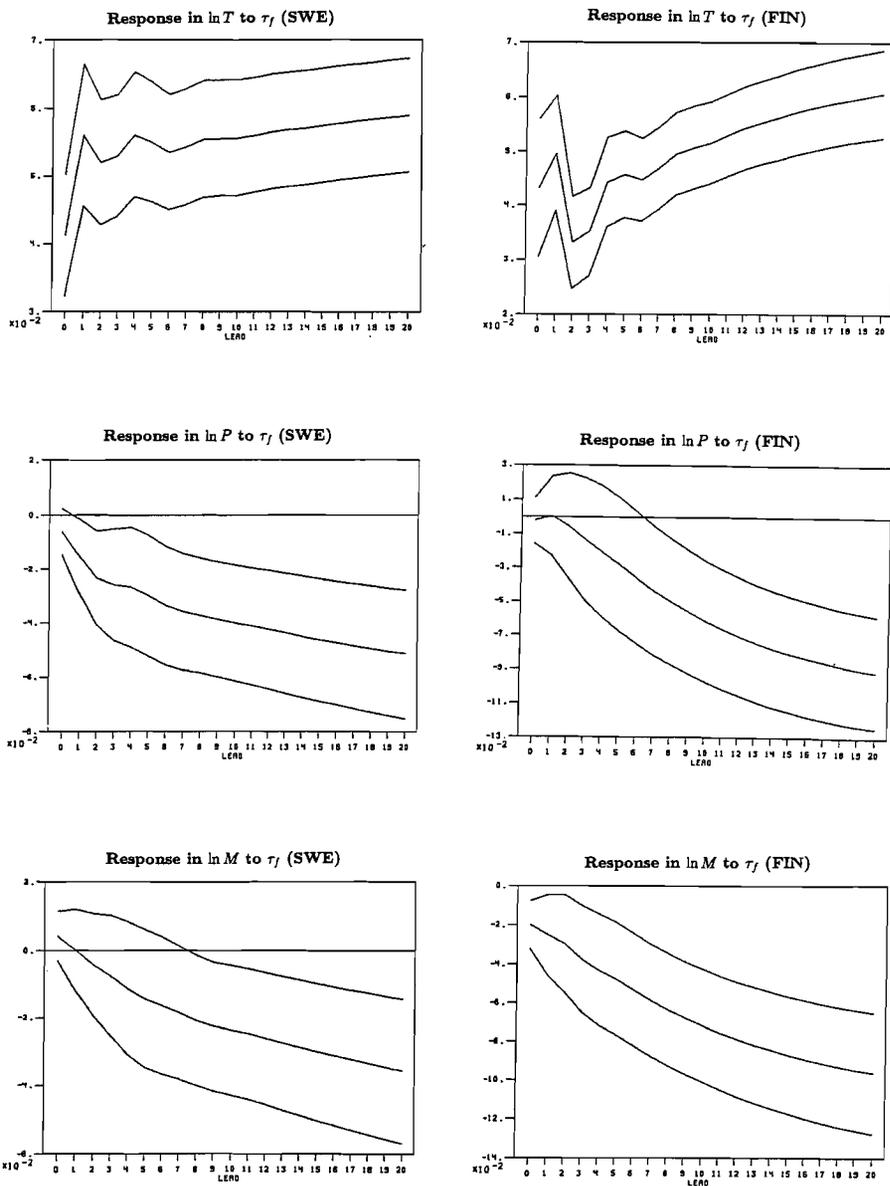
**Finland:**

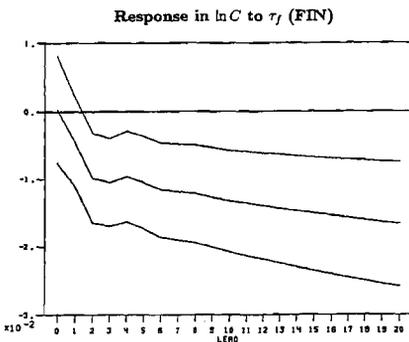
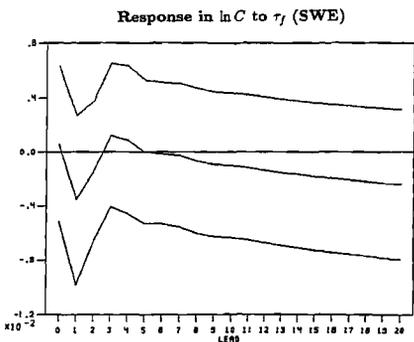
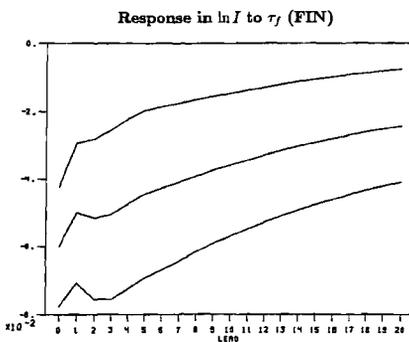
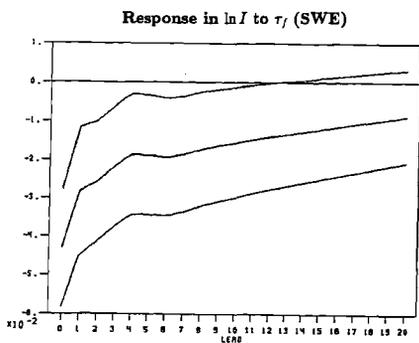
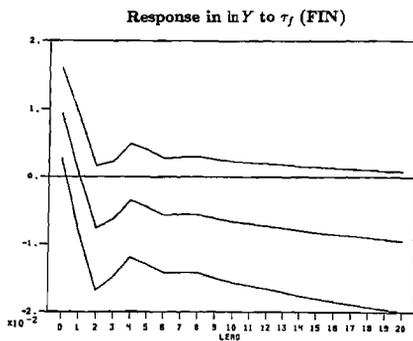
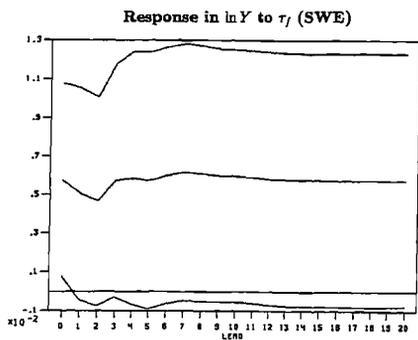
$$\begin{bmatrix} \ln T_t \\ \ln P_t \\ \ln M_t \\ \ln Y_t \\ \ln I_t \\ \ln C_t \end{bmatrix} = x_0 + \begin{bmatrix} .0657 & .0 & .0 \\ (.0043) & (-) & (-) \\ -.0943 & -.0167 & .1414 \\ (.0145) & (.0131) & (.0092) \\ -.1079 & .0495 & .1414 \\ (.0156) & (.0135) & (.0092) \\ -.0135 & .0663 & .0 \\ (.0062) & (.0043) & (-) \\ -.0135 & .0663 & .0 \\ (.0062) & (.0043) & (-) \\ -.0223 & .0643 & .0 \\ (.0061) & (.0042) & (-) \end{bmatrix} \begin{bmatrix} \hat{\tau}_{f,t} \\ \hat{\tau}_{r,t} \\ \hat{\tau}_{n,t} \end{bmatrix} + \hat{u}_t.$$

**Note:** The estimated asymptotically normal standard errors within parenthesis are based on Theorem 2.3. The vector  $u_t$  is equal to  $\tilde{C}(L)\varepsilon_t$ .

In section 3.3 it was noted that theoretical arguments suggest that there should be three cointegrating vectors for each economy. The unrestricted maximum likelihood estimates of these three vectors are also given in Table 3.2. In addition, I have imposed the restrictions on the first two cointegrating vectors implied by the conditions that the log of the velocity of money and the log of the investments–output ratio are stationary. Although these restrictions are strongly rejected by the likelihood ratio tests, it can be seen that *if* these constraints are imposed on the data generating process, then the estimated coefficients of the third cointegrating vector are consistent with the theoretical restriction that the coefficients on the money stock and the price level should sum to zero. Moreover, they are very similar in the two countries. Since the unconstrained estimates of the three cointegrating vectors are quite similar (cf. Table 3.2) this is perhaps not so surprising. Furthermore, King, Plosser, Stock and Watson [53] use the log of the ratio between consumption and output as their third cointegration vector, which substantially differs from the ones obtained here.

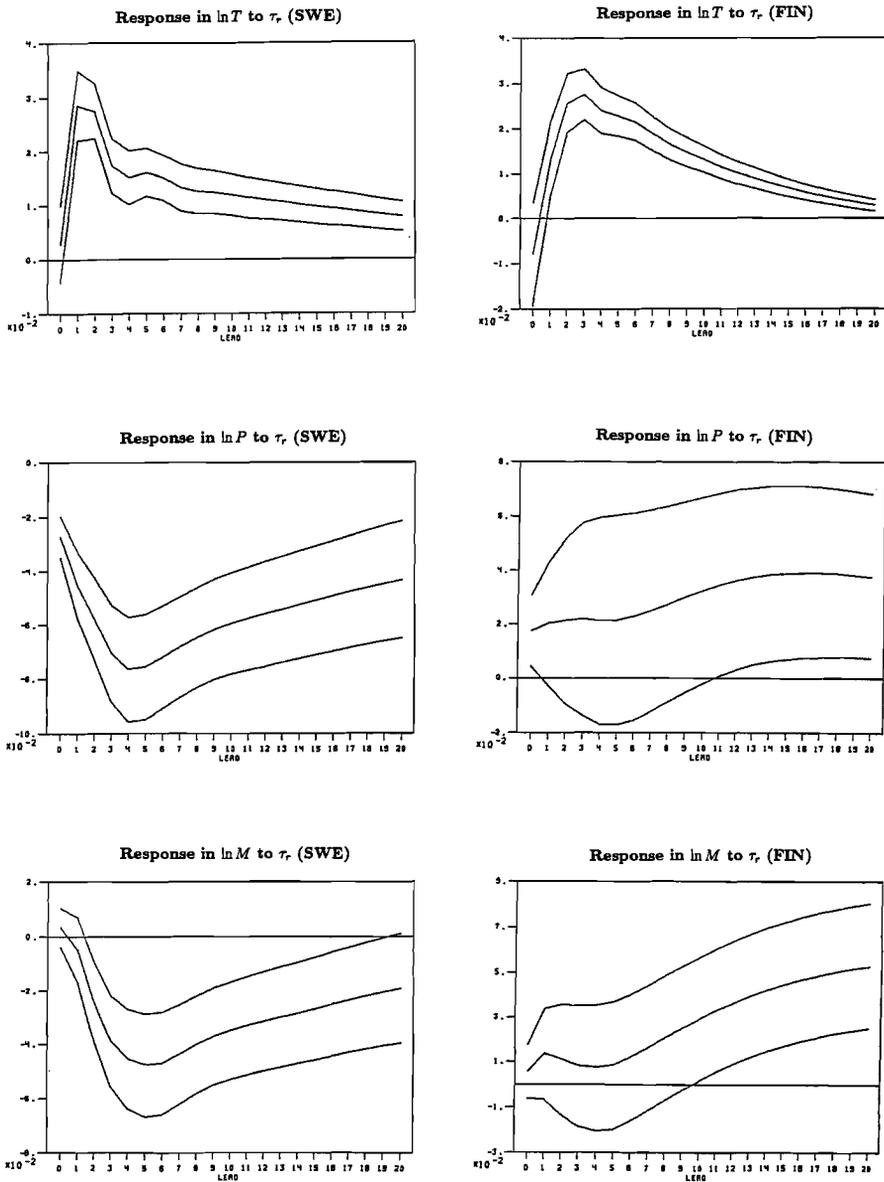
Figure 3.1: Responses in  $x_t$  to a one standard deviation shock to the foreign trend ( $\tau_{f,t}$ ) for Sweden (SWE) and Finland (FIN) with 95 percent confidence bounds.



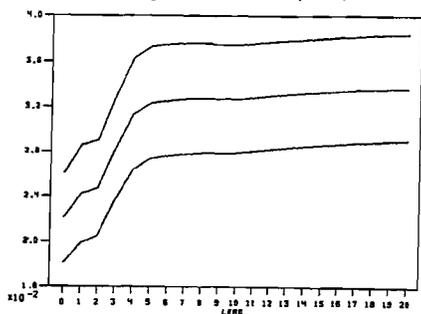


**Note:** The estimated asymptotically normal standard errors from which the confidence bounds have been computed are based on Theorem 2.3.

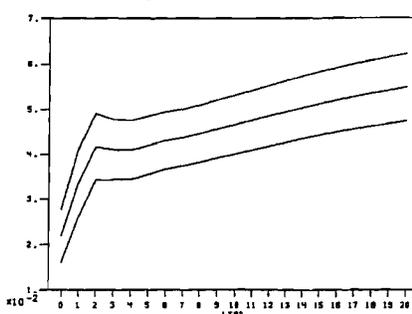
Figure 3.2: Responses in  $x_t$  to a one standard deviation shock to the real domestic trend ( $\tau_{r,t}$ ) for Sweden (SWE) and Finland (FIN) with 95 percent confidence bounds.



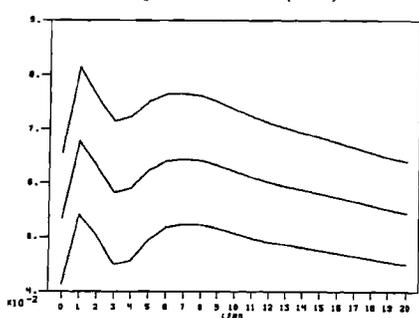
Response in  $\ln Y$  to  $\tau_r$  (SWE)



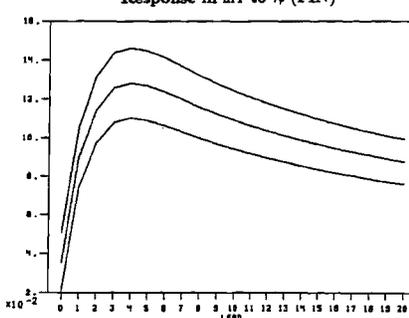
Response in  $\ln Y$  to  $\tau_r$  (FIN)



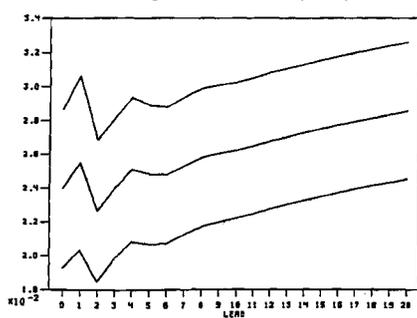
Response in  $\ln I$  to  $\tau_r$  (SWE)



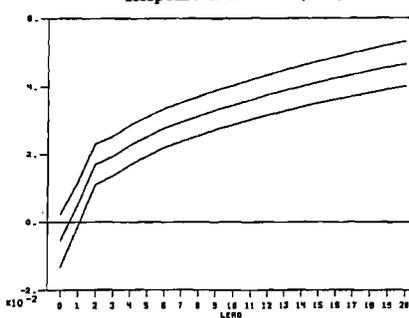
Response in  $\ln I$  to  $\tau_r$  (FIN)



Response in  $\ln C$  to  $\tau_r$  (SWE)

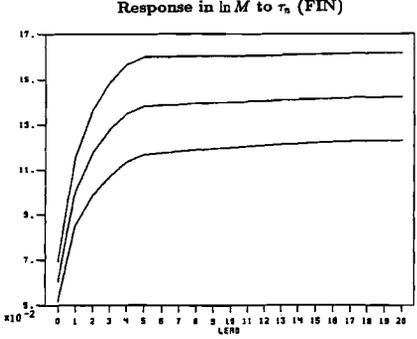
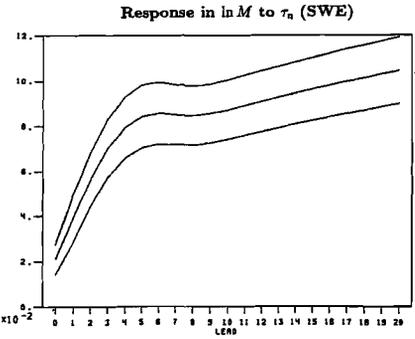
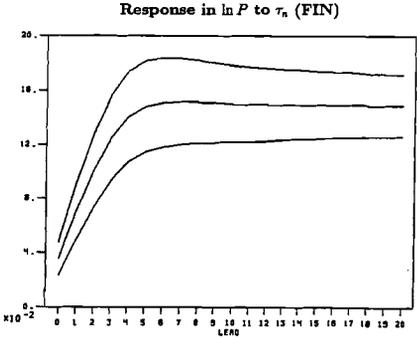
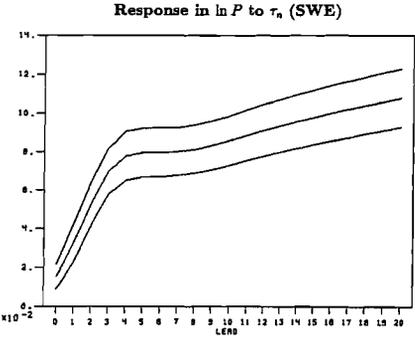
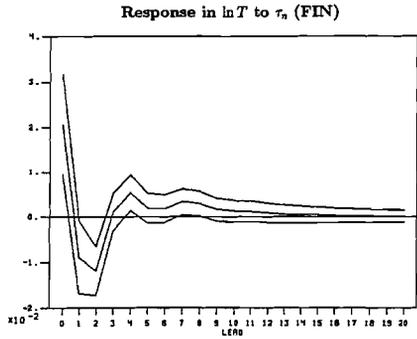
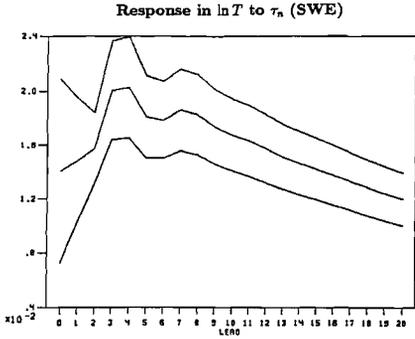


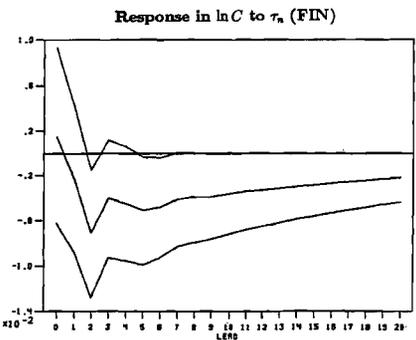
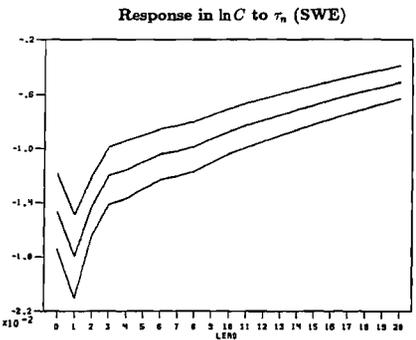
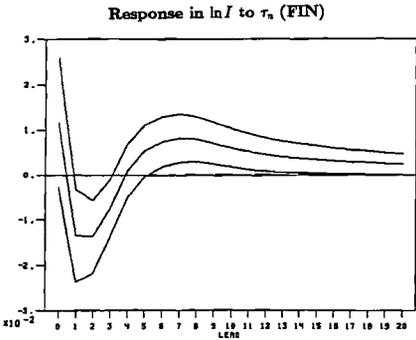
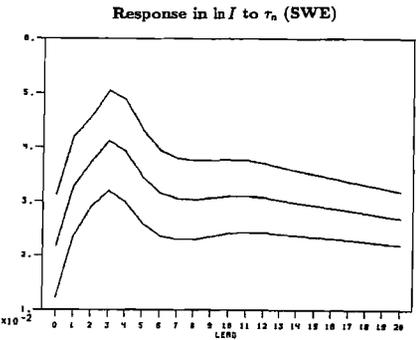
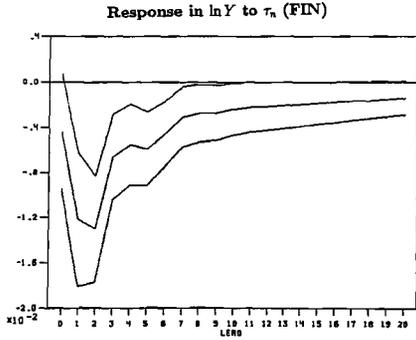
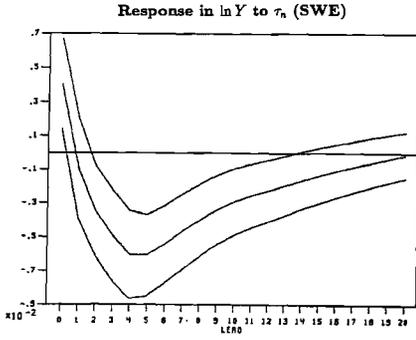
Response in  $\ln C$  to  $\tau_r$  (FIN)



**Note:** The estimated asymptotically normal standard errors from which the confidence bounds have been computed are based on Theorem 2.3.

Figure 3.3: Responses in  $x_t$  to a one standard deviation shock to the nominal domestic trend ( $\tau_{n,t}$ ) for Sweden (SWE) and Finland (FIN) with 95 percent confidence bounds.





Note: The estimated asymptotically normal standard errors from which the confidence bounds have been computed are based on Theorem 2.3.

Hence, the pretesting of data suggests that (i) a simple vector autoregressive model may not be a good description of the time series, and (ii) conditional on such a model there is basically no support for the economic steady state relationships. It is possible that this reflects difficulties in measuring the variables with a “reasonable degree of accuracy” and/or the different economic conditions that have been prevalent during, e.g., the pre and post world wars periods. I shall nevertheless continue the illustration since these results do not stand in conflict with the purpose of this chapter.

In Table 3.3 the estimated common trends models are presented. Recall from section 3.3 that the  $A = (A_0\pi)$  matrix has been chosen so that the first element of  $\tau_t$  is interpreted as a foreign trend, while the second and third elements are interpreted as real and nominal domestic trends, respectively. These interpretations are based on the restrictions of both countries being small open economies (domestic shocks have zero long run effect on terms of trade, as reflected in the two zero elements in the first row of  $\hat{A}$ ) and of long run neutrality of money with respect to effects on the real variables (innovations to the nominal trend have zero long run impact on real variables).

The parameters of the estimated common trends models for the two countries are very similar. Apart from the coefficients on the foreign trend in the output and investments equations, the signs of the estimated parameters are equal in the two countries. In most cases, the coefficients for the Finnish model are somewhat greater (in absolute terms) than those for the Swedish model. It should be noted that the standard errors for estimated parameters are based on treating  $\alpha$  and  $A_0$  as known. Yet, *two* of the parameters in  $\alpha$  and  $A_0$  are not constrained.<sup>10</sup> This, however, does not seem overly problematic.

The responses (in log-levels) to innovations to the stochastic trends are displayed in Figures 3.1 – 3.3 along with 95 percent confidence bounds. A major impression from a quick glance at these figures is that the responses are very much the same for the two countries. Looking first at the effects from shocks to the foreign trend (Figure 3.1), the only important difference is that the response of Finnish consumption is negative, while the effects on Swedish consumption are not “significantly” different from zero.

In Figure 3.2 the response functions from innovations to the real domestic trend are depicted. The restriction that the long run response of terms of trade should be zero does not seem to be “seriously in conflict” with the data. The maximum response is recorded after about two years. Hence, within a business cycle horizon, a positive real domestic shock tends to improve terms of trade and then vanish. Also, the long run responses in the price level and the money stock are negative and positive, respectively, for both countries (cf. Table 3.3), whereas the adjustment in the short and medium term is quite different and typically very slow.

The response functions for an innovation to the nominal domestic trend are shown in Figure 3.3. In terms of the responses of output and consumption, the restriction of long run neutrality does not seem to be “in conflict” the data sets neither for Finland nor for Sweden. The responses in terms of trade and investments in Sweden, however, are rather far from the theoretical predictions. In Finland, the responses of these variables satisfy the theoretical restrictions almost exactly at the business cycle horizon.

---

<sup>10</sup>It is misleading here to count the number of unknown parameters of  $\alpha$  and  $A_0$  in (3.14) and (3.15). If we premultiply  $\alpha'$  in (3.14) by the inverse of the matrix obtained from the final 3 columns of  $\alpha'$  and compare the parameters of the transformed  $\alpha$  matrix to those in  $A_0$ , it becomes clear that there are only two free parameters in these two matrices.

Table 3.4: Ratio of  $s$  steps ahead forecast error variance of  $x_t$  which is accounted for by shocks to the foreign trend ( $\tau_{f,t}$ ) for Sweden (S) and Finland (F).

$k$	$\ln T_t$		$\ln P_t$		$\ln M_t$		$\ln Y_t$		$\ln I_t$		$\ln C_t$	
	S	F	S	F	S	F	S	F	S	F	S	F
1	.546 (.063)	.328 (.071)	.018 (.025)	.001 (.006)	.012 (.021)	.086 (.050)	.046 (.038)	.066 (.044)	.248 (.070)	.329 (.071)	.000 (.004)	.000 (.002)
2	.622 (.052)	.423 (.064)	.036 (.034)	.000 (.001)	.003 (.006)	.059 (.042)	.037 (.034)	.025 (.019)	.167 (.061)	.247 (.064)	.006 (.009)	.006 (.009)
3	.653 (.047)	.421 (.060)	.049 (.039)	.001 (.004)	.003 (.003)	.053 (.040)	.033 (.033)	.024 (.007)	.143 (.058)	.207 (.061)	.005 (.009)	.026 (.020)
4	.679 (.045)	.432 (.058)	.052 (.040)	.002 (.008)	.004 (.008)	.057 (.041)	.033 (.032)	.023 (.009)	.124 (.054)	.182 (.059)	.005 (.006)	.040 (.027)
5	.703 (.043)	.469 (.056)	.051 (.040)	.005 (.011)	.007 (.012)	.062 (.043)	.032 (.032)	.019 (.010)	.110 (.052)	.166 (.057)	.004 (.003)	.045 (.030)
6	.721 (.041)	.503 (.055)	.054 (.041)	.008 (.014)	.010 (.016)	.068 (.045)	.031 (.031)	.018 (.011)	.100 (.050)	.154 (.056)	.003 (.003)	.050 (.033)
7	.735 (.039)	.529 (.053)	.059 (.042)	.012 (.018)	.013 (.018)	.076 (.046)	.030 (.031)	.017 (.013)	.094 (.049)	.146 (.055)	.003 (.003)	.056 (.036)
8	.746 (.038)	.556 (.051)	.064 (.044)	.018 (.021)	.016 (.021)	.085 (.048)	.030 (.031)	.017 (.014)	.090 (.048)	.140 (.054)	.003 (.002)	.061 (.037)
9	.757 (.037)	.583 (.049)	.070 (.045)	.023 (.024)	.019 (.023)	.094 (.050)	.030 (.031)	.017 (.015)	.085 (.047)	.136 (.054)	.002 (.003)	.065 (.039)
10	.766 (.036)	.609 (.047)	.075 (.047)	.030 (.027)	.023 (.025)	.104 (.052)	.030 (.031)	.017 (.016)	.082 (.046)	.132 (.054)	.002 (.003)	.068 (.040)
15	.805 (.031)	.712 (.038)	.095 (.051)	.065 (.038)	.038 (.033)	.148 (.058)	.030 (.031)	.018 (.020)	.069 (.043)	.120 (.053)	.002 (.005)	.082 (.045)
20	.833 (.027)	.782 (.030)	.111 (.054)	.099 (.046)	.053 (.038)	.182 (.062)	.029 (.031)	.021 (.023)	.060 (.040)	.112 (.052)	.003 (.007)	.090 (.047)
$\infty$	1.00 (-)	1.00 (-)	.156 (.063)	.305 (.071)	.124 (.058)	.341 (.071)	.067 (.046)	.040 (.036)	.067 (.046)	.040 (.036)	.009 (.018)	.107 (.054)

Note: The estimated standard errors within parenthesis are based on Theorem 2.4.

Let us now turn to the questions about the relative importance of the various disturbances that were mentioned in the introduction of the chapter. These questions can be approached through the decompositions of forecast error variances which are reported in Tables 3.4 – 3.7.

**Foreign versus domestic shocks:** In both countries, shocks to the foreign trend account for around 20 percent of the forecast error variance of investments in the short run, while their role is negligible in the long run. In the cases of output and consumption, the role of permanent foreign shocks is negligible in the short run as well. Permanent real shocks, on the other hand, account for most of the variance in output at the four year horizon (70 percent in Sweden and 60 percent in Finland; cf. Table 3.5). They also account for a substantial proportion of the variance in investments and, for Sweden, in the price level. These findings may suggest that terms of trade is not a good proxy for investigating foreign influence on small open economies. Instead or as a complement, a measure on foreign output seems to be a plausible alternative. The foreign trend may then be interpreted as reflecting, e.g., technology “import”. That, however, requires a different economic model for deriving the cointegration vectors.

Table 3.5: Ratio of  $s$  steps ahead forecast error variance of  $x_t$  which is accounted for by shocks to the real domestic trend ( $\tau_{r,t}$ ) for Sweden (S) and Finland (F).

$k$	$\ln T_t$		$\ln P_t$		$\ln M_t$		$\ln Y_t$		$\ln I_t$		$\ln C_t$	
	S	F	S	F	S	F	S	F	S	F	S	F
1	.002 (.006)	.011 (.016)	.362 (.072)	.060 (.042)	.008 (.017)	.007 (.015)	.673 (.054)	.358 (.070)	.383 (.066)	.113 (.045)	.638 (.054)	.017 (.024)
2	.106 (.026)	.024 (.007)	.367 (.071)	.034 (.032)	.006 (.006)	.013 (.020)	.677 (.052)	.453 (.066)	.470 (.065)	.381 (.058)	.616 (.056)	.018 (.007)
3	.137 (.029)	.070 (.018)	.374 (.069)	.024 (.027)	.053 (.032)	.010 (.017)	.696 (.050)	.553 (.058)	.494 (.064)	.531 (.059)	.634 (.054)	.079 (.024)
4	.124 (.027)	.109 (.024)	.389 (.068)	.019 (.024)	.103 (.047)	.007 (.014)	.717 (.047)	.613 (.053)	.490 (.064)	.620 (.057)	.657 (.051)	.127 (.036)
5	.110 (.026)	.123 (.026)	.397 (.068)	.016 (.022)	.134 (.054)	.006 (.013)	.740 (.045)	.653 (.049)	.490 (.064)	.673 (.056)	.676 (.049)	.175 (.047)
6	.104 (.025)	.130 (.027)	.396 (.068)	.015 (.021)	.151 (.057)	.005 (.012)	.759 (.043)	.685 (.046)	.501 (.064)	.706 (.055)	.692 (.047)	.221 (.050)
7	.099 (.024)	.135 (.028)	.391 (.069)	.014 (.021)	.160 (.059)	.005 (.012)	.774 (.042)	.712 (.043)	.515 (.063)	.728 (.054)	.706 (.045)	.265 (.054)
8	.093 (.023)	.134 (.028)	.383 (.069)	.014 (.021)	.163 (.060)	.006 (.013)	.785 (.041)	.736 (.041)	.529 (.063)	.744 (.053)	.718 (.044)	.306 (.056)
9	.088 (.022)	.130 (.027)	.374 (.069)	.015 (.021)	.162 (.060)	.007 (.014)	.795 (.040)	.756 (.039)	.540 (.062)	.756 (.053)	.730 (.043)	.344 (.058)
10	.083 (.022)	.125 (.026)	.363 (.069)	.016 (.022)	.158 (.060)	.009 (.016)	.804 (.039)	.773 (.037)	.550 (.062)	.765 (.053)	.740 (.041)	.379 (.059)
15	.067 (.018)	.096 (.021)	.308 (.068)	.022 (.025)	.132 (.056)	.024 (.023)	.836 (.036)	.833 (.031)	.582 (.060)	.797 (.052)	.785 (.035)	.520 (.058)
20	.056 (.016)	.072 (.016)	.258 (.065)	.027 (.027)	.106 (.050)	.040 (.028)	.858 (.035)	.868 (.029)	.606 (.058)	.816 (.051)	.818 (.030)	.616 (.055)
$\infty$	.0 (-)	.0 (-)	.010 (.018)	.010 (.015)	.008 (.015)	.007 (.037)	.933 (.046)	.960 (.036)	.933 (.046)	.960 (.036)	.991 (.018)	.893 (.054)

Note: The estimated standard errors within parenthesis are based on Theorem 2.4.

**Real versus nominal shocks:** As can be seen from Table 3.6, the influence of shocks to the nominal trend on fluctuations in real variables is small, except for the cases of consumption and investments in Sweden. The most striking results in Tables 3.5 and 3.6 are related to consumption. In Finland, a rather small fraction of the short run forecast error variance of consumption is accounted for by shocks to the real domestic trend. The role of the nominal trend is practically zero (like it is for the other real variables). In the Swedish data, most of the variance in consumption (65 percent at the four year horizon) is accounted for by shocks to the real domestic trend, but nominal shocks are important as well (25 percent at the four year horizon). Another interesting observation that can be made from Table 3.5 is associated with investments. In Finland, shocks to the real domestic trend initially have a quite small impact on the forecast error variance of investments and then rapidly becomes more and more important. For Sweden we find that such shocks account for about 40 percent of the forecast error variance in investments at the one year horizon and slowly become more important. Also, permanent real shocks are unimportant for the nominal domestic variables in Finland in the short and medium run, while this is not the case in Sweden.

Table 3.6: Ratio of  $s$  steps ahead forecast error variance of  $x_t$  which is accounted for by shocks to the nominal domestic trend ( $\tau_{n,t}$ ) for Sweden (S) and Finland (F).

$k$	$\ln T_t$		$\ln P_t$		$\ln M_t$		$\ln Y_t$		$\ln I_t$		$\ln C_t$	
	S	F	S	F	S	F	S	F	S	F	S	F
1	.064 (.030)	.075 (.038)	.113 (.044)	.242 (.067)	.298 (.072)	.816 (.051)	.023 (.015)	.014 (.016)	.063 (.027)	.012 (.015)	.238 (.046)	.001 (.007)
2	.054 (.020)	.049 (.017)	.173 (.050)	.296 (.070)	.364 (.071)	.804 (.047)	.011 (.005)	.047 (.026)	.098 (.031)	.013 (.004)	.270 (.048)	.002 (.002)
3	.058 (.017)	.050 (.013)	.253 (.054)	.338 (.070)	.438 (.067)	.778 (.047)	.012 (.002)	.055 (.025)	.127 (.034)	.012 (.004)	.270 (.047)	.013 (.010)
4	.070 (.017)	.042 (.011)	.316 (.055)	.370 (.070)	.491 (.063)	.762 (.048)	.015 (.005)	.046 (.021)	.153 (.037)	.009 (.003)	.251 (.045)	.013 (.011)
5	.077 (.018)	.037 (.010)	.354 (.057)	.399 (.069)	.527 (.062)	.756 (.049)	.019 (.007)	.040 (.019)	.166 (.039)	.007 (.002)	.234 (.042)	.013 (.012)
6	.079 (.018)	.032 (.008)	.377 (.058)	.422 (.068)	.553 (.061)	.754 (.050)	.021 (.008)	.036 (.017)	.167 (.039)	.006 (.002)	.221 (.041)	.014 (.013)
7	.081 (.018)	.028 (.007)	.392 (.059)	.441 (.068)	.573 (.061)	.753 (.051)	.022 (.008)	.032 (.016)	.164 (.039)	.005 (.002)	.211 (.039)	.014 (.013)
8	.083 (.018)	.026 (.007)	.406 (.060)	.456 (.067)	.589 (.061)	.751 (.052)	.021 (.008)	.028 (.014)	.162 (.039)	.005 (.001)	.201 (.038)	.014 (.013)
9	.084 (.018)	.024 (.006)	.419 (.061)	.470 (.066)	.603 (.061)	.749 (.053)	.020 (.008)	.025 (.013)	.160 (.038)	.005 (.001)	.192 (.036)	.013 (.012)
10	.084 (.018)	.022 (.006)	.432 (.062)	.481 (.065)	.616 (.061)	.746 (.054)	.019 (.008)	.023 (.012)	.159 (.038)	.005 (.001)	.184 (.035)	.012 (.012)
15	.079 (.016)	.015 (.004)	.493 (.064)	.519 (.062)	.675 (.060)	.722 (.059)	.014 (.006)	.015 (.008)	.162 (.038)	.004 (.001)	.148 (.030)	.010 (.010)
20	.071 (.015)	.011 (.003)	.545 (.066)	.542 (.061)	.718 (.058)	.696 (.062)	.010 (.005)	.011 (.006)	.165 (.037)	.004 (.001)	.122 (.025)	.008 (.007)
$\infty$	.0 (-)	.0 (-)	.834 (.064)	.685 (.071)	.869 (.059)	.587 (.070)	.0 (-)	.0 (-)	.0 (-)	.0 (-)	.0 (-)	.0 (-)

Note: The estimated standard errors within parenthesis are based on Theorem 2.4.

At a forecast horizon longer than three years, around 16 percent of the variance in Swedish investments is accounted for by shocks to the nominal trend. On average, however, permanent nominal shocks do not appear to be very important. Another example is provided by the behavior of the price level. At the four year horizon, only about 35 percent of its variance (in both countries) is accounted for by shocks to the nominal trend. Finally, shocks to the nominal trend account for about 80 percent of the forecast error variance of the money stock in Finland at the one to four year horizon but only for around 40 percent in Sweden. In the long run this picture is reversed.

**Permanent versus transitory shocks:** In a common trends framework, the most interesting results when it comes to innovation accounting are those related to the importance of permanent shocks in relation to transitory shocks. In particular, the question of how important growth innovations are within a business cycle perspective can be addressed. It can be noted that once the cointegration vectors have been determined, the choice of  $A_0$  is irrelevant for the interpretation of the permanent and transitory shocks as groups of shocks. That is, the small open economy and long run nominal neutrality assumptions do not influence *any* figures reported in Table 3.7.

Table 3.7: Ratio of  $s$  steps ahead forecast error variance of  $x_t$  which is accounted for by the three permanent shocks for Sweden (S) and Finland (F).

$k$	$\ln T_t$		$\ln P_t$		$\ln M_t$		$\ln Y_t$		$\ln I_t$		$\ln C_t$	
	S	F	S	F	S	F	S	F	S	F	S	F
1	.612 (.057)	.414 (.070)	.393 (.067)	.303 (.071)	.318 (.072)	.909 (.016)	.742 (.042)	.438 (.069)	.694 (.048)	.454 (.068)	.921 (.022)	.018 (.025)
2	.782 (.034)	.496 (.060)	.576 (.060)	.330 (.071)	.373 (.071)	.876 (.021)	.725 (.044)	.525 (.062)	.735 (.041)	.641 (.049)	.892 (.018)	.026 (.012)
3	.848 (.024)	.541 (.054)	.676 (.048)	.373 (.070)	.494 (.064)	.841 (.026)	.741 (.041)	.632 (.051)	.764 (.037)	.750 (.036)	.909 (.016)	.118 (.032)
4	.873 (.020)	.583 (.050)	.757 (.038)	.391 (.070)	.598 (.056)	.826 (.029)	.765 (.038)	.682 (.046)	.767 (.036)	.811 (.028)	.913 (.015)	.180 (.044)
5	.890 (.018)	.629 (.046)	.802 (.031)	.420 (.069)	.668 (.048)	.824 (.029)	.791 (.034)	.712 (.043)	.766 (.036)	.846 (.023)	.914 (.015)	.233 (.051)
6	.904 (.016)	.665 (.043)	.827 (.028)	.445 (.068)	.714 (.043)	.827 (.029)	.818 (.031)	.739 (.040)	.768 (.036)	.866 (.020)	.916 (.014)	.285 (.056)
7	.915 (.014)	.692 (.040)	.842 (.026)	.467 (.067)	.746 (.039)	.834 (.028)	.826 (.029)	.761 (.037)	.773 (.035)	.879 (.018)	.920 (.014)	.335 (.058)
8	.922 (.013)	.716 (.038)	.853 (.024)	.488 (.065)	.768 (.036)	.842 (.026)	.836 (.027)	.781 (.034)	.781 (.034)	.889 (.016)	.922 (.013)	.381 (.059)
9	.929 (.012)	.737 (.036)	.863 (.022)	.508 (.064)	.784 (.034)	.850 (.025)	.845 (.026)	.798 (.032)	.785 (.034)	.897 (.015)	.924 (.013)	.422 (.059)
10	.933 (.011)	.756 (.034)	.870 (.021)	.527 (.062)	.797 (.032)	.859 (.024)	.853 (.025)	.813 (.030)	.791 (.033)	.902 (.014)	.926 (.013)	.459 (.058)
15	.951 (.008)	.823 (.026)	.896 (.017)	.606 (.055)	.845 (.025)	.894 (.018)	.880 (.021)	.866 (.022)	.813 (.030)	.921 (.012)	.935 (.011)	.612 (.051)
20	.960 (.007)	.865 (.020)	.914 (.014)	.668 (.048)	.877 (.020)	.918 (.014)	.897 (.018)	.900 (.017)	.831 (.028)	.932 (.010)	.943 (.010)	.714 (.041)
$\infty$	1.00 (-)											

Note: The estimated standard errors within parenthesis are based on Corollary 2.2.

I have already stressed the role of permanent real domestic shocks to the fluctuations of output and investments in Finland and Sweden. Table 3.7 summarizes the role of all three permanent shocks.<sup>11</sup> There, we find that these innovations generally account for most (about 60 percent) of the fluctuations at the business cycle (four year) horizon in both countries. Transitory shocks, as a group, exert important influence only on the price level (60 percent) and consumption (80 percent) in Finland. It has been noted above that the role of permanent foreign and nominal shocks is rather small. In principle, transitory foreign and nominal shocks could still be of importance. However, considering the great influence of permanent shocks, it may be concluded that transitory foreign and transitory nominal shocks *potentially* can be important only to the price level and consumption in Finland.

<sup>11</sup>It can be noted that the forecast error variance decompositions of the transitory shocks as a group can be derived immediately from Table 3.7 by multiplying a figure by  $-1$  and adding 1. The standard errors of these figures are identical to those for the permanent shocks. It can be noted that all standard errors reported above are such that the closer the variance decomposition is to 0 or 1, the smaller is the error (cf. section 2.4).

### 3.5 Concluding Remarks

In this chapter I have illustrated how the common trends approach can be implemented in practise. It should probably be emphasized again that whether the conclusions reached from the analysis of impulse response functions and forecast error variance decompositions above should be regarded as “empirical regularities” of the Swedish and the Finnish economies has neither been addressed nor been my purpose to discuss. Instead I have focused on the basic ideas in common trends modelling and how economic theory can be made use of. Furthermore, some of the practical problems has been examined, e.g., lag order selection and tests of the theoretical cointegration vectors. For the Finnish and Swedish samples it is found from such analysis of the common trends model that data are not well described by these simple “structures”.

I have not mentioned any “informal tests” of the cointegration vectors. It is, however, straightforward to compute estimates of the transitory components from the restricted vector autoregression (cf. Warne and Vredin [100]). For example, graphs of these components may suggest whether theoretical cointegration vectors (which have been rejected by the LR test) are completely unreasonable or not. For the Swedish and the Finnish data I have used above such graphs strongly suggest that estimates of the transitory components are unlikely to be well modelled as wide sense stationary. Hence, the trends are misspecified as well. Although such “tests” of the cointegration vectors are statistically unsatisfactory they may at least shed some light on how well the trends are accounted for and thus act as a complement to the formal tests.

# Chapter 4

## Stationarity, Cointegration and a Rational Expectations Model of the Term Structure

### 4.1 Introduction

One of the basic problems in monetary and financial economics is to explain how returns on bonds with different time to maturity are related. From the point of view of conducting and predicting monetary policy it is often of great importance for the monetary authority as well as for investors in financial markets to know how and to what extent changes in, e.g., short term rates influence longer term rates. A popular family of models for explaining the behavior in the term structure of interest rates is what is generally known as expectations models (cf. Cox, Ingersoll and Ross [18]). A common denominator for these models is that the interest rate on a long term bond can be expressed as a function of current and expected future short term rates.

Hypotheses which follow from simple expectations models have generally been rejected by data (see, e.g., Campbell and Shiller [14], Fama [30,31], and Mankiw and Summers [66]), although there are some exceptions (Froot [32] and Sargent [81]). In most of these papers interest rates are transformed by means of first differences and cointegration relationships (cf. Engle and Granger [26]) prior to testing the economic hypothesis. The argument for using such transformations is often that standard asymptotic theory is based on stationary data and that deviations from such a property need to be accounted for. Consequently, there seems to be a consensus among empirical researchers that interest rates are not stationary<sup>1</sup> and that nonstationarity probably biases inference in models where this has not been dealt with.

When data are generated by a linear time series model, first differences and cointegra-

---

<sup>1</sup>Ekdahl [24] and Rose [78], for example, test for unit roots in interest rate series. In both studies the conclusion is that nominal interest rates for Sweden and the U.S., respectively, are nonstationary in levels but stationary in first differences. However, given that the Fisher relationship holds, their evidence on the hypothesis that real interest rates are stationary differ; while Rose does not reject nonstationary real interest rates, Ekdahl finds that Swedish real interest rates may be stationary. The reason for this difference appears to be that Rose only considers an *approximation* of the Fisher hypothesis, while Ekdahl also uses the exact formulae.

tion relationships are linear restrictions on the data generating process. If such restrictions are not supported by data,<sup>2</sup> then tests of economic hypotheses as well as estimated parameters are asymptotically biased. On the other hand, if nonstationarity is not accounted for, inference is generally based on incorrect limiting distributions (cf. Sims, Stock and Watson [92]). This is, of course, a standard dilemma in empirical analyses. Nevertheless, this type of problem is rarely examined in practise.

In this chapter I shall analyse and test a linear rational expectations model of the term structure of interest rates on Swedish time series data. The maintained hypothesis is that data can be represented by a finite order vector autoregression. In connection with inference I shall consider some fundamental statistical questions. First, I will examine if it matters whether data is modelled as jointly stationary or cointegrated of order (1,1). Second, the procedure for testing the economic hypothesis involves a nonlinear reparameterization of the original constraints. From Gregory and Veall [35] it is well known that the commonly applied Wald test is *not* numerically invariant in finite samples with respect to such reparameterizations. It therefore seems natural to compare the performance of the Wald test with a test statistic which is invariant, e.g., the likelihood ratio test. Third, the limiting  $\chi^2$  distribution is not necessarily a good approximation of the small sample uncertainty of these two statistics. Hence, empirical distributions of the tests are interesting alternatives to the  $\chi^2$ . I will approach these three questions mainly by means of bootstrap resampling (cf. Efron [23]).

The remainder of the chapter is structured as follows. In section 4.2 I discuss the term structure model. Section 4.3 is concerned with connecting the data generating process to the economic and nonstationarity restrictions. Specifically, I shall examine implications from the term structure model on the data generating process and how we can test for cointegration and the term structure model jointly. To my knowledge, no such test has previously been considered. The empirical results are presented in section 4.4, and conclusions are given in section 4.5. In the appendix I show that standard inference is valid in an unrestricted vector autoregression if interest rates are cointegrated of order (1,1) with a stationary spread. The analysis there relies on the results obtained by Sims, Stock and Watson [92].

## 4.2 An Expectations Model of the Term Structure

If there is no uncertainty about future bond prices, a no arbitrage condition generally implies that the returns on holding different bonds, say, one period are equal regardless of time to maturity. However, this is not necessarily true when future bond prices are uncertain. Instead, agents may require a term premium to invest in a bond with a time to maturity different from their investment horizon. In that case, expected holding period yields are equal across maturities except for a term premium. In this chapter, I shall test whether the term premium for a long term bond in relation to a short term bill is constant over time when investors are assumed to have rational expectations.

---

<sup>2</sup>For example, time series are jointly stationary *or* nonstationarity is not reflected in terms of linear trends in the first and second moments but rather from some other (nonexplosive) functions of time. Furthermore, models which incorporate conditional heteroskedasticity, such as the GARCH and GARCH-M models, may be more suitable descriptions of the data generating process (cf. Bollerslev [11] and Engle, Lilien and Robins [27]).

### 4.2.1 The Local Expectations Hypothesis

Let  $H_{t+1}^{(\tau)}$  denote the one-period ahead holding period yield on a bond which matures at time  $t + \tau$ . Furthermore,  $r_t$  denotes the time  $t$  yield to maturity on a one-period pure discount bond, and  $E$  the mathematical expectations operator. In the terminology of Cox, Ingersoll and Ross [18], the *local expectations hypothesis* states that

$$E[H_{t+1}^{(\tau)} | \mathcal{A}_t] = r_t + \psi^{(\tau)}. \quad (4.1)$$

Here,  $\mathcal{A}_t$  is the information set which agents use at  $t$ , and  $\psi^{(\tau)}$  is a constant term premium, associated with  $\tau$ -period bonds.

An often overlooked advantage from considering the local expectations hypothesis is that it is consistent with coupon carrying bonds. In contrast, the yield to maturity, return to maturity, and unbiasedness expectations hypotheses (cf. Cox, Ingersoll and Ross [18]) are stated in terms of interest rates on discount bonds. There are, at least, two reasons why this is problematic. First, trade in or even existence of any long term pure discount bonds is very rare in actual markets. Second, interest rates on bonds carrying coupons generally provide poor approximations to interest rates on pure discount bonds (cf. Litzenberger and Rolfo [56]). Hence, it is convenient to examine a hypothesis which allows for coupon carrying bonds when one analyses the long term part of the maturity spectrum.

In section 4.4 I shall treat the long term bond as a consol. To simplify the discussion about the local expectations hypothesis I will do the same here. Thus, let  $R_t$  denote yield to maturity measured as a one-period rate,  $H_{t+1}$  the one-period ahead holding period yield,  $\psi$  the term premium, and  $C$  the coupon for a consol. It is assumed that coupons are paid once per period, prior to trade, i.e., if an investor purchases a consol at  $t$ , (s)he will receive the first coupon at  $t + 1$ , the second at  $t + 2$  if (s)he does not sell the consol at  $t + 1$ , and so on. The time  $t$  price of a consol,  $P_t$ , is related to yield to maturity by

$$P_t = \sum_{i=1}^{\infty} \frac{C}{(1 + R_t)^i} = \frac{C}{R_t}, \quad (4.2)$$

where  $R_t > 0$  ensures that  $P_t$  is finite and with  $C > 0$  we have that  $P_t$  is positive. The one-period ahead holding period yield on the consol is therefore

$$H_{t+1} := \frac{P_{t+1} + C}{P_t} - 1 = \frac{R_t(1 + R_{t+1})}{R_{t+1}} - 1, \quad (4.3)$$

by substituting for  $P_t$  and  $P_{t+1}$  from (4.2). From this relationship between holding period yield and yield to maturity it can be seen that if  $R_t$  and  $R_{t+1}$  are observable, so is  $H_{t+1}$ , i.e., information on the value of the coupon is not needed. For a  $\tau$ -period coupon carrying bond paying a principal of, e.g., \$1.00 this is no longer true (cf. Shiller [89]).

To obtain a *linear* relationship between  $R_t$  and  $r_t$  from (4.1) and (4.3) we must linearize the latter equation. A first order Taylor expansion of  $H_{t+1}$  about  $E[R_t] = \bar{R}$  and some algebra give us

$$H_{t+1} \approx \frac{1}{1 - \rho} R_t - \frac{\rho}{1 - \rho} R_{t+1}, \quad (4.4)$$

where  $\rho := 1/(1 + \bar{R})$ . Substituting the approximation of  $H_{t+1}$  in (4.1) and assuming that  $R_t \in \mathcal{A}_t$ , we find that the local expectations hypothesis can be written as

$$E[R_{t+1} | \mathcal{A}_t] = \frac{1}{\rho} R_t + \frac{\rho - 1}{\rho} (r_t + \psi). \quad (4.5)$$

Subtracting  $R_t$  from both sides of (4.5) we obtain the version of the local expectations hypothesis which is tested by Mankiw and Summers [66]. Ruling out rational bubbles,<sup>3</sup> the forward solution to the linear stochastic difference equation in (4.5) is

$$R_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i \mathbb{E}[r_{t+i} | \mathcal{A}_t] + \psi. \quad (4.6)$$

Hence, the local expectations hypothesis and the linear approximation of the one-period ahead holding period yield jointly imply that the consol yield can be expressed as a weighted average of current and expected future short term rates plus a term premium. The weights on the short term rates sum to one and are exponentially declining in the distance between  $t$  and  $t + i$ , i.e., short term rates in the immediate future are assigned a larger weight than short term rates which are realized at a later date.

Equation (4.6) is, in fact, a special case of Shiller's [89] linear rational expectations model of the term structure of interest rates for  $\tau$ -period coupon carrying bonds. As shown by Campbell and Shiller [14,15], an equivalent form of (4.6) is

$$R_t - r_t = \sum_{i=1}^{\infty} \rho^i \mathbb{E}[r_{t+i} - r_{t+i-1} | \mathcal{A}_t] + \psi. \quad (4.7)$$

That is, the spread between the long term and the short term yield to maturity is a weighted average of expected future changes in the short term yield plus a term premium.

## 4.2.2 Discussion of the Model

Present value models such as (4.6) and (4.7) have been examined for stock prices and dividends as well as the long term part of the maturity spectrum for bonds (see, e.g., Baillie [4] and Campbell and Shiller [14]). Whether one decides to test (4.6) or (4.7) generally depends on ones prior views about which statistical assumptions seem to comply best with data. In particular, if interest rates are modelled as jointly stationary, it is straightforward to test (4.6) while (4.7) is less appropriate. On the other hand, if interest rates are modelled as nonstationary in levels but stationary in first differences, Campbell and Shiller [14,15] argue that (4.7) is better suited for the exercise. Their argument is based on substituting the conditional expectation of the first difference of  $r_{t+i}$  for the realization of the variable and a stationary, purely nondeterministic forecast error. Since  $0 < \rho < 1$ , it follows that the spread,  $R_t - r_t$ , is stationary. In the terminology of Engle and Granger [26], interest rates are cointegrated of order (1,1) with one cointegration vector.

As noted above, from equation (4.3) it can be seen that  $H_{t+1}$  can be directly inferred from observations on  $R_t$  and  $R_{t+1}$ . Since the local expectations hypothesis is a linear (affine) relationship between the expected one-period ahead holding period yield and the current short term rate, it seems reasonable to ask why a linearization is considered. Clearly, if we would like to examine a linear term structure relationship between the short and the long term yield a linearization is required. If the error from using the approximation in (4.4) instead of the exact formulae in (4.3) is small, then inference should not be very different when analysing a linear instead of a nonlinear model. Campbell [12]

---

<sup>3</sup>See Campbell and Shiller [14, page 1066] for a discussion on this matter.

has shown that linear approximations such as (4.4) are often very good for long term bonds. In terms of the long term yield series I shall study in section 4.4, the estimated mean error between  $H_{t+1}$  in (4.3) and (4.4) is  $-.00077$ , when measured as a monthly holding period yield. Furthermore, the estimated error variance is  $.000019$ , the ratio between the estimated error variance and the estimated variance of  $H_{t+1}$  in (4.3) is  $.012$ , while the estimated correlation between the two measures of holding period yield is  $.994$ . Hence, the approximation seems to provide quite accurate values for the holding period yield series.<sup>4</sup>

Also, the Taylor expansion is based on the assumption that the long term rate is not “too variable” (see, e.g., Campbell [12]). If interest rates are modelled as nonstationary in levels and stationary in first differences, terms involving second order effects and higher in (4.4) may have a tendency to diverge from zero. Accordingly, the approximation may eventually become very unreliable. To examine this possibility, suppose we compare  $(H_{t+1} - r_t)$  to its first order Taylor expansion about  $\bar{R} - \bar{r}$ . My reason for looking at such an expansion is simply to see if (4.7) can be derived from the local expectations hypothesis when interest rates are nonstationary in levels but not in first differences. If  $\bar{R}$  and  $\bar{r}$  exist, such an approximation and the local expectations hypothesis imply that

$$E[R_{t+1} - r_{t+1} | \mathcal{A}_t] = \frac{1}{\rho} (R_r - r_t) - \frac{1}{\rho} E[r_{t+1} - r_t | \mathcal{A}_t] + \frac{\rho - 1}{\rho} \psi. \quad (4.8)$$

Ruling out rational bubbles, the forward solution to (4.8) is given by the right hand side in (4.7). Thus, the approximation in (4.4) may still serve its purpose even if higher moments of the long term rate are explosive.

## 4.3 Connecting Economic Theory with Data

### 4.3.1 The Data Generating Model

In most empirical studies on the term structure it is assumed that interest rates are either stationary in levels or after differencing once. Although the latter assumption is by far more common than the former, I shall model the data generating mechanism such that one of the cases is true. The argument for allowing for both possibilities is simply that it is virtually impossible to statistically distinguish many nonstationary stochastic processes from a stationary process.

Let  $x_t := [r_t \ R_t]'$  and suppose that the bivariate stochastic process  $\{x_t\}_{t=1}^{\infty}$  is generated according to

$$A(L)x_t = \mu + \varepsilon_t. \quad (4.9)$$

Here,  $A(L)$  is a  $2 \times 2$  matrix polynomial of order  $p$  in the lag operator,  $L$ , i.e.,

$$A(L) := I_2 - \sum_{k=1}^p A_k L^k,$$

$I_2$  is the  $2 \times 2$  identity matrix,  $L^k x_t = x_{t-k}$ , and for all  $k \in \{1, \dots, p\}$

$$A_k = \begin{bmatrix} a_{11k} & a_{12k} \\ a_{21k} & a_{22k} \end{bmatrix}.$$

---

<sup>4</sup>Graphs on the two holding period yield series are available from the author on request.

It is assumed that  $\det[A(\lambda)] \neq 0$  for all  $|\lambda| < 1$ , i.e.,  $\{x_t\}_{t=1}^{\infty}$  is not an explosive process. Furthermore,  $\mu = [\mu_r \ \mu_R]'$  is a vector of constants which is related to the unconditional mean of  $x_t$ , while  $\varepsilon_t = [\varepsilon_{r,t} \ \varepsilon_{R,t}]'$ , and the innovations  $\{\varepsilon_t\}_{t=1}^{\infty}$  are independent and identically distributed with mean zero and positive definite covariance matrix  $\Sigma$ . Finally,  $\{x_0, \dots, x_{1-p}\}$  are  $2p$  purely deterministic variables which serve as initial conditions to  $x_t$  in (4.9).

If  $\{x_t\}_{t=1}^{\infty}$  is jointly wide sense stationary, by the Wold Decomposition Theorem (see, e.g., Hannan [38, page 137]) we know that the matrix polynomial  $A(L)$  is invertible, i.e.,  $\det[A(\lambda)] \neq 0$  for all  $|\lambda| \leq 1$ . On the other hand, if  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with the spread between the long and the short term yield being stationary, from Campbell and Shiller [14,15] and Theorem 2.1 we find that an *equivalent* representation of the data generating process in (4.9) is

$$B(L)y_t = \theta + \eta_t. \tag{4.10}$$

Here, we may let  $y_t := [\Delta r_t \ (R_t - r_t)]'$ ,  $\Delta := 1 - L$  is the first difference operator,  $\theta := [\mu_r \ (\mu_R - \mu_r)]'$ ,  $\eta_t := [\varepsilon_{r,t} \ (\varepsilon_{R,t} - \varepsilon_{r,t})]'$ , and  $E[\eta_t \eta_t'] = \Omega$ . Hence,  $\{y_t\}_{t=1}^{\infty}$  is jointly wide sense stationary and the matrix polynomial  $B(L) := I_2 - \sum_{k=1}^p B_k L^k$  is invertible. Furthermore, the relationship between  $A(L)$  and  $B(L)$  is given by

$$A_k = \begin{cases} M^{-1}(B_1 + D_{\perp})M & \text{if } k = 1, \\ M^{-1}(B_k - B_{k-1}D_{\perp})M & \text{if } k \in \{2, \dots, p\}, \end{cases} \tag{4.11}$$

where

$$M := \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad D_{\perp} := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

and  $B_p D_{\perp} = 0$ . It can be noted that  $y_t = (I_2 - D_{\perp}L)Mx_t$  and that the second row of  $M$ , denoted by  $m_2$ , corresponds to the cointegration vector. Also,  $D_{\perp}$  is specified such that the first difference operator is applied to  $m_1 x_t$ , but not to  $m_2 x_t$ . Details on the connection between  $A(L)$  and  $B(L)$  are given in Theorem 2.1.

A few observations can be made on the basis of equation (4.9) and (4.10). First, if interest rates are cointegrated of order (1,1) with one cointegration vector, the matrix  $A(1)$  has rank 1 and has one unit root. To see this, note that

$$A(1) = M^{-1}B(1)(I_2 - D_{\perp})M.$$

Here,  $M$  and  $B(1)$  both have rank 2, while  $(I_2 - D_{\perp})$  has rank 1.<sup>5</sup> Accordingly, since  $A(1)$  is of reduced rank, it follows that the matrix polynomial  $A(L)$  cannot be inverted with standard techniques and that cointegration imposes the set of linear restrictions in (4.11) on the data generating process in (4.9). Second, if interest rates are jointly wide sense stationary, the matrix  $A(1)$  has rank 2 since  $A(L)$  is invertible, i.e., the data generating process for  $x_t$  is misspecified in equation (4.10).

---

<sup>5</sup>It can be noted that if  $A(1)$  has rank 1, then  $A(\lambda)$  does not necessarily have *one* unit root. However, if  $\{x_t\}_{t=1}^{\infty}$  is nonstationary in levels and jointly wide sense stationary in first differences, it follows that  $\text{rank}[A(1)] = 1$  implies that  $A(\lambda)$  has one unit root. On the other hand, if  $A(\lambda)$  has, say, two unit roots when  $A(1)$  has rank 1, then interest rates are nonstationary in first differences but not in second differences.

Since the representation in equation (4.9) is valid for interest rates which are jointly wide sense stationary as well as cointegrated of order (1,1), one may ask why one even bothers about the representation in (4.10). A *brief* comment on this nontrivial question may be presented in the following way. If we estimate the vector autoregression in (4.9) by, e.g., multivariate least squares and interest rates are nonstationary, the estimated regression coefficients will generally not have a joint asymptotic normal distribution (cf. Sims, Stock and Watson [92]). Furthermore, statistics such as the Wald and the likelihood ratio (LR) tests generally do not have a limiting  $\chi^2$  distribution. Supposing that the nonstationarity of interest rates can be modelled by the restrictions in equation (4.11), then estimates of the regression coefficients in the transformed model (4.10) and of the Wald and the LR tests have their standard limiting properties. Hence, inference can be conducted in (4.10) with standard procedures.

However, there are some cases when the limiting distributions of the regression coefficients in (4.9) and the corresponding test statistics are standard even if the time series are nonstationary. According to Sims, Stock and Watson [92, page 114]:

... estimators of coefficients in the original untransformed model have a joint nondegenerate asymptotic normal distribution if the model can be rewritten so that these original coefficients correspond in the transformed model to coefficients on mean zero stationary canonical regressors ... when all the restrictions being tested in the untransformed model correspond to restrictions on the coefficients of mean zero stationary canonical regressors in the transformed model, then the test statistic [Wald test; my remark] has the usual limiting  $\chi^2$  distribution ...

In terms of the data generating process being studied here, the untransformed model is given in (4.9). The transformed model considered by Sims, Stock and Watson, however, does not correspond to that in (4.10). In the appendix I derive a transformed model of the Sims, Stock and Watson type and show that *all* coefficients in (4.9) correspond to coefficients on mean zero stationary canonical regressors in the transformed model *when* interest rates are cointegrated of order (1,1) with a stationary spread. Also, the restrictions I test in section 4.4 satisfy the conditions which ensure that the Wald (and LR) test has a limiting  $\chi^2$  distribution (see also West [101]).

Despite the fact that standard inference from statistical analysis of (4.9) is valid when interest rates are cointegrated of order (1,1) with a stationary spread, it is an open question whether tests should be conducted from estimated coefficients of this representation or from estimated coefficients of equation (4.10). As noted by Sims, Stock and Watson [92], tests based on the two versions of the data generating process may be very different, e.g., in terms of small sample accuracy and degree of pretest bias. I shall address such issues to some extent in section 4.4. In particular, I will examine if it matters greatly for inference whether one tests the term structure model in (4.9) or (4.10) when one does *not* know if the latter represents data correctly.

### 4.3.2 Term Structure Restrictions

In section 4.2 I derived two equivalent term structure relationships for the long term and the short term yield to maturity. To connect the economic hypothesis with the data generating model in section 4.3.1, let us assume that

- (i)  $\mathcal{A}_t$ , the information set which agents use at  $t$ , includes at least current and lagged values of  $x_t$  for all  $t \geq 1$ , and
- (ii)  $E[x_{t+i}|\mathcal{A}_t] = P[x_{t+i}|\mathcal{A}_t]$  for all  $t, i \geq 1$ .

Here,  $P[x_{t+i}|\mathcal{A}_t]$  denotes a linear projection of  $x_{t+i}$  on the set  $\mathcal{A}_t$ . Given that agents' expectations are modelled as rational and symmetric, these two assumptions do not seem very restrictive when data is generated by a linear time series model.

Let  $\mathcal{E}_t := \{x_{t-k}\}_{k=0}^{t+p-1}$  be the set of time  $t$  information which is available to an econometrician. By assumption (i) it is found that  $\mathcal{E}_t \subset \mathcal{A}_t$  for all  $t \geq 1$ . This result and assumption (ii) imply that

$$P[P[x_{t+i}|\mathcal{A}_t]|\mathcal{E}_t] = P[x_{t+i}|\mathcal{E}_t],$$

i.e., the law of iterated projections applies to agents' expectations. Since  $R_t = P[R_t|\mathcal{E}_t]$ , a linear projection of both sides of equation (4.6) on  $\mathcal{E}_t$  gives us

$$R_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i P[r_{t+i}|\mathcal{E}_t] + \psi. \quad (4.12)$$

In contrast to equation (4.6), this hypothesis can be tested. Specifically, the linear projections on the right hand side can be derived from the data generating process in (4.9), whereas agents' conditional expectations are unknown in the present framework.

To derive a linear projection of  $r_{t+i}$  on  $\mathcal{E}_t$ , let us rewrite (4.9) as a first order vector autoregression. We then get

$$\begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_2 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & 0 & & I_2 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

or

$$X_t = U + AX_{t-1} + E_t. \quad (4.13)$$

Here,  $X_t$ ,  $U$ , and  $E_t$  are  $2p \times 1$  vectors and  $A$  is a  $2p \times 2p$  matrix.

Let the  $2 \times 2p$  matrix  $J_p := [I_2 \ 0 \cdots 0]$  so that  $x_t = J_p X_t$ , whereas  $U = J_p' \mu$  and  $E_t = J_p' \varepsilon_t$ . Now, by recursive substitution for  $X_{t+i-1}, \dots, X_{t+1}$  in equation (4.13), we find that  $X_{t+i}$  may be written as

$$X_{t+i} = \sum_{j=0}^{i-1} A^j U + A^i X_t + \sum_{j=0}^{i-1} A^j E_{t+i-j}.$$

Premultiplying both sides by  $J_p$  and using the functions for  $U$  and  $E_t$ , we obtain

$$x_{t+i} = \sum_{j=0}^{i-1} J_p A^j J_p' \mu + J_p A^i X_t + \sum_{j=0}^{i-1} J_p A^j J_p' \varepsilon_{t+i-j}. \quad (4.14)$$

Projecting both sides of (4.14) on  $\mathcal{E}_t$ , we have that

$$P[x_{t+i}|\mathcal{E}_t] = \sum_{j=0}^{i-1} J_p A^j J_p' \mu + J_p A^i X_t, \quad (4.15)$$

for all  $i \geq 1$ . Here, the forecast error is given by  $\sum_{j=0}^{i-1} J_p A^j J_p' \varepsilon_{t+i-j}$ , a moving average process of order  $i - 1$ .

Let  $x_t$  be measured about its unconditional mean so that  $\mu$  and  $\psi$  can be cancelled.<sup>6</sup> Also, let  $N_0 := [0 \ 1]$  and  $N_1 := [1 \ 0]$  so that  $R_t = N_0 x_t = N_0 J_p X_t$  and  $P[r_{t+i} | \mathcal{E}_t] = N_1 P[x_{t+i} | \mathcal{E}_t]$  for all  $t, i \geq 1$ . Substituting these expressions into (4.12) and using (4.15), we get

$$N_0 J_p X_t = (1 - \rho) N_1 J_p \sum_{i=0}^{\infty} (\rho A)^i X_t.$$

This relationship must hold for any realisation of  $X_t$  when the vector autoregression in (4.9) is consistent with the term structure model in (4.12). Accordingly, we find that

$$N_0 J_p = (1 - \rho) N_1 J_p \sum_{i=0}^{\infty} (\rho A)^i = (1 - \rho) N_1 J_p (I_{2p} - \rho A)^{-1}. \quad (4.16)$$

The second equality is implied by the assumptions that  $R_t > 0$  (i.e.,  $\bar{R} > 0$ ) and that  $\{x_t\}_{t=1}^{\infty}$  is not an explosive stochastic process.<sup>7</sup>

Postmultiplying both sides in (4.16) by  $(I_{2p} - \rho A)$  and rearranging terms, we obtain the following set of term structure restrictions on the data generating process in (4.9):

$$[0 \ \rho] \begin{bmatrix} A_1 & \cdots & A_p \end{bmatrix} = [(\rho - 1) \ 1] \begin{bmatrix} I_2 & 0 & \cdots & 0 \end{bmatrix}. \quad (4.17)$$

Since the parameters  $\{A_k\}_{k=1}^p$  are unknown, equation (4.17) suggests that  $\rho$  should be known prior to testing the term structure constraints. Campbell and Shiller [14] use an estimate of  $\bar{R}$  to calculate  $\rho$ . Alternatively, one can, and according to Shea [87] should, replace  $\bar{R}$  with the mean of the coupon. For the long term yield series I study in section 4.4 all estimated parameters and test statistics are equal for these two estimates of  $\rho$ . Hence, I shall only report results for an estimate of  $\rho$  which is based on  $\bar{R}$ .

In case we model interest rates as cointegrated of order (1,1), it can be shown that for the vector autoregression in (4.10) to be consistent with the term structure model in (4.7), given assumptions (i) and (ii) above, the following linear cross equation restrictions must hold:<sup>8</sup>

$$[\rho \ \rho] \begin{bmatrix} B_1 & \cdots & B_p \end{bmatrix} = [0 \ 1] \begin{bmatrix} I_2 & 0 & \cdots & 0 \end{bmatrix}. \quad (4.18)$$

A quick glance at the restrictions in (4.17) and (4.18) may suggest that they are algebraically different. This is, however, not the case.

From equation (4.11) we have that there is a (unique) mapping from the parameter space of the  $B_k$  matrices to the parameter space of the  $A_k$  matrices. Imposing the term

<sup>6</sup>From equations (4.6) and (4.7), it can be shown that  $E[R_t] - E[r_t] = \psi$ . Hence, measuring the yield series about their unconditional means implies that  $\psi$  must be excluded from the term structure model. Similarly, taking expectations of both sides in (4.9) and letting  $E[x_t] = 0$  for all  $t$ , it follows that  $\mu$  should be excluded.

<sup>7</sup>Mathematically, if  $\{x_t\}_{t=1}^{\infty}$  is not an explosive time series, then  $\det[A(\lambda)] \neq 0$  for all  $|\lambda| < 1$ , or equivalently, the eigenvalues of  $A$  are on or inside the unit circle. Consequently, with  $\bar{R} > 0$ , the eigenvalues of  $\rho A$  are inside the unit circle. Since all eigenvalues of  $I_{2p}$  are unity, it follows that  $(I_{2p} - \rho A)$  is nonsingular.

<sup>8</sup>For a derivation, see Campbell and Shiller [14].

structure restrictions in (4.17) on the  $A_k$  matrices, we find that

$$A_k = \begin{cases} \begin{bmatrix} a_{111} & a_{121} \\ (\rho - 1)/\rho & 1/\rho \end{bmatrix} & \text{if } k = 1, \\ \begin{bmatrix} a_{11k} & a_{12k} \\ 0 & 0 \end{bmatrix} & \text{if } k \in \{2, \dots, p\}. \end{cases} \quad (4.19)$$

Similarly, the term structure restrictions in (4.18) imply that

$$B_k = \begin{cases} \begin{bmatrix} b_{111} & b_{121} \\ -b_{111} & 1/\rho - b_{121} \end{bmatrix} & \text{if } k = 1, \\ \begin{bmatrix} b_{11k} & b_{12k} \\ -b_{11k} & -b_{12k} \end{bmatrix} & \text{if } k \in \{2, \dots, p\}. \end{cases} \quad (4.20)$$

If we apply the nonstationarity restrictions in (4.11) to these  $B_k$  matrices we get

$$A_k = \begin{cases} \begin{bmatrix} 1 + b_{111} - b_{121} & b_{121} \\ (\rho - 1)/\rho & 1/\rho \end{bmatrix} & \text{if } k = 1, \\ \begin{bmatrix} b_{11,k} - b_{11,k-1} - b_{12,k} & b_{12,k} \\ 0 & 0 \end{bmatrix} & \text{if } k \in \{2, \dots, p\}. \end{cases} \quad (4.21)$$

Hence, the term structure restrictions in (4.17) and (4.18) are algebraically equivalent if interest rates are cointegrated of order (1,1) with a stationary spread. I shall make use of equations (4.19)–(4.21) below in relation to discussions on Granger-causality and a simple and intuitive explanation as to why jointly wide sense stationary interest rates can be *nearly* cointegrated.

Before we turn to that, it should be noted that these term structure restrictions have a simple economic interpretation. The constraints in (4.19) suggest that future excess returns on long term bonds are unpredictable given information on current and lagged  $r_t$  and  $R_t$ . A far greater problem is that of interpreting a statistical rejection of the restrictions. Campbell and Shiller [14] propose an approach based on replacing  $A$  in (4.16) with an unconstrained estimate and apply these parameters on  $X_t$ . This gives us a theoretical but unconstrained estimate of the right hand side in (4.6) which can be compared to the historical  $R_t$  time series. I shall use this approach when the need occurs.

### 4.3.3 Implications for Statistical Analysis

In order to test the constraints in (4.17) and (4.18), classical asymptotic theory provides us with at least three standard test statistics; the Wald, the LR, and the Lagrange Multiplier (LM) tests.<sup>9</sup> When selecting between these statistics, researchers tend to pick the one which is the easiest to apply. Although such a selection procedure is not always recommendable, the properties of the tests are often very similar.

However, the constraints in (4.17) and (4.18) are derived from *nonlinear* reparameterizations of the original nonlinear cross equation restrictions (see, e.g., the step from

---

<sup>9</sup>For a survey, see Engle [25], and for analyses on implicit alternatives and asymptotic local power, see Davidson and MacKinnon [19].

equation (4.16) to (4.17)). The study of Gregory and Veall [35] shows that the Wald test is *not* numerically invariant in finite samples to nonlinear transformations of algebraically equivalent restrictions and that this can substantially effect size and power of the Wald test. Furthermore, the numerical noninvariance follows from the observation that the Wald test is based on a first order Taylor expansion of the constraints. Accordingly, algebraically equivalent nonlinear restrictions lead to nontrivial differences in the corresponding first order approximations. In contrast, the LR and LM tests are not based on Taylor series expansions and consequently not subject to such numerical noninvariance. For this reason I shall examine the Wald and the LR tests in the empirical analyses to see whether the choice of a test statistic appears to be important. The method I will apply is bootstrap resampling which, under somewhat mild assumptions, gives us fairly accurate estimates of the unknown small sample distributions (cf. Efron [23]).

Second, the term structure model has some interesting implications for Granger-causality. From the constraints in (4.19) it can be seen that  $r_{t-1}$  Granger-causes  $R_t$ . While this is trivial, a more interesting proposition is the following:

**Proposition 4.1** *If  $\{R_{t-s}\}_{s=1}^p$  does not Granger-cause  $r_t$ , the restrictions in (4.19) are satisfied by the data generating process in (4.9), and the term structure model in (4.6) holds true, then  $\{R_t\}_{t=1}^\infty$  and  $\{r_t\}_{t=1}^\infty$  are both explosive stochastic processes.*

It may appear as if the "... and the term structure model ..." condition would be redundant. However, this is, as will be discussed below, not the case. Similarly,

**Proposition 4.2** *If  $\{x_t\}_{t=1}^\infty$  is generated by (4.9) such that  $\det[A(\lambda)] \neq 0$  for all  $|\lambda| < 1$  and satisfies the restrictions in (4.19), then  $\{R_{t-s}\}_{s=1}^p$  Granger-causes  $r_t$ .*

An informal proof of these propositions will be given below. Before we come to that, note that if  $\{x_t\}_{t=1}^\infty$  is an explosive stochastic process, so is  $\{y_t\}_{t=1}^\infty$ .

Let  $a_{ij}(\lambda)$  denote the  $(i, j)$ :th element of  $A(\lambda)$ . The determinant of the matrix polynomial  $A(\lambda)$  is then given by

$$\det[A(\lambda)] = a_{11}(\lambda)a_{22}(\lambda) - a_{12}(\lambda)a_{21}(\lambda). \quad (4.22)$$

The economic restrictions in (4.19) implies that  $a_{21}(\lambda) = -(\rho - 1)\rho^{-1}\lambda$  and  $a_{22}(\lambda) = (1 - \rho^{-1}\lambda)$ . Accordingly, the determinant of  $A(\lambda)$  conditional on these scalar polynomials is

$$\det[A(\lambda)] = a_{11}(\lambda)(1 - \rho^{-1}\lambda) + a_{12}(\lambda)(\rho - 1)\rho^{-1}\lambda. \quad (4.23)$$

Suppose  $\{R_{t-s}\}_{s=1}^p$  does not Granger-cause  $r_t$ , i.e.,  $a_{12}(\lambda) = 0$ . A root to  $\det[A(\lambda)] = 0$  can then be obtained from  $(1 - \rho^{-1}\lambda) = 0$ , i.e.,  $\lambda = \rho$  and  $\lambda$  is consequently inside the unit circle since the discount factor  $\rho$  takes values in the open unit interval. Hence,  $\{R_t\}_{t=1}^\infty$  is explosive. Furthermore, from equation (4.6) it can be seen that an explosive long term yield series is not compatible with a nonexplosive short term yield series.<sup>10</sup> Accordingly, a test of the hypothesis  $a_{12k} = 0$  for all  $k \in \{1, \dots, p\}$  conditional on the term structure

---

<sup>10</sup>Alternatively, if  $a_{12k} = 0$  for all  $k \in \{1, \dots, p\}$ , then  $A$  has an eigenvalue equal to  $\rho^{-1}$  and  $\rho A$  therefore a unit eigenvalue. Hence,  $(I_{2p} - \rho A)$  is singular and  $\lim_{i \rightarrow \infty} (\rho A)^i \neq 0$ . It then follows that the derivation of the term structure constraints from (4.12) to (4.16) is invalid. The same conclusions apply to similar expressions for  $y_t$ .

model is based on explosive time series and standard inference (if any) does therefore not apply. Also, if we test this hypothesis without conditioning on the term structure model, it follows that a true null *cannot* be consistent with the term structure model if we model interest rates as a joint nonexplosive stochastic process. Hence, if the term structure model holds true and the corresponding restrictions in (4.17) are supported by the nonexplosive data generating process in (4.9), it must follow that  $a_{12k} \neq 0$  for some  $k \in \{1, \dots, p\}$ , i.e.,  $\{R_{t-s}\}_{s=1}^p$  Granger-causes  $r_t$ .

In relation to a proposition in Campbell and Shiller [15, page 513] an interesting observation can be made on the basis of equation (4.20). The Campbell and Shiller proposition states that: If the cointegrated present value model in (4.7) holds, then *either*  $(R_t - r_t)$  is an exact linear function of current and lagged  $\Delta r_t$  or  $\{(R_{t-s} - r_{t-s})\}_{s=1}^p$  Granger-causes  $\Delta r_t$  (and  $\Delta R_t$ ). In a vector autoregressive framework with all roots on and outside the unit circle, there is no doubt about the validity of this statement. However, what Campbell and Shiller fail to see is that the “either” statement can be dispensed with. From (4.20), i.e., from the cointegrated vector autoregressive parameters which are conditioned on the present value model in (4.7), it can be seen that if  $(R_t - r_t)$  (second row of the  $B_k$  matrices) is an exact linear function of lagged  $\Delta r_t$ , then  $b_{121} = 1/\rho$  and  $b_{12k} = 0$  for all  $k \in \{2, \dots, p\}$ . Accordingly,  $(R_{t-1} - r_{t-1})$  is associated with a nonzero coefficient in the  $\Delta r_t$ -equation. Hence, the “or” statement is implied. In fact, the proposition in Campbell and Shiller is a special case of Proposition 4.2 above.

From a practical perspective, Granger–noncausality tests are only meaningful if the parameters in the data generating process are not conditioned on the term structure model. Furthermore, Propositions 4.1 and 4.2 suggest that a failure to reject Granger–noncausality in the  $r_t$ -equation should make us suspicious of the validity of the term structure model. On the other hand, a failure to reject Granger–noncausality in the  $R_t$ -equation is less of a problem, especially if  $\rho$  is very close to 1, i.e.,  $a_{211} = (\rho - 1)/\rho \approx 0$ , while  $a_{21k} = 0$  for all  $k \geq 2$  when the term structure model is consistent with data.

Moreover,  $a_{221} = 1/\rho \approx 1$  suggests that the long term rate *almost* has a unit root. Since Proposition 4.2 implies that  $a_{12k} \neq 0$  for some  $k$ , it follows that the short term rate almost has a unit root as well. Thus, even if interest rates are modelled as jointly wide sense stationary, the term structure model and  $\rho \approx 1$  implies that interest rates are nearly cointegrated. This may be expressed more formally as:

**Proposition 4.3** *If  $\{x_t\}_{t=1}^\infty$  is generated by (4.9) and satisfies the restrictions in (4.19), then  $\lim_{\rho \rightarrow 1} \det[A(1)] = 0$ .*

The result follows immediately from noting that  $\det[A(1)] = [a_{11}(1) + a_{12}(1)](\rho - 1)/\rho$ . In practise we expect  $\rho$  to be close to one for high frequency data. It may thus be virtually impossible to distinguish wide sense stationary from cointegrated interest rates when data are (nearly) consistent with the term structure model.

### 4.3.4 A Joint Test of Cointegration and the Term Structure Model

In view of the fact that an estimator (least squares) of the parameters in equation (4.9) has a joint nondegenerate asymptotic normal distribution when interest rates are cointegrated of order (1,1) with a stationary spread, I shall propose a likelihood ratio test which allows

me to test the cointegration and the economic theory restrictions *jointly*. Moreover, the test statistic has a limiting  $\chi^2$  distribution. To my knowledge, a similar test has never been considered previously in the cointegration literature. To make the analysis more general, I shall first introduce some new notation which I relate to the rational expectations model studied here.

Let the  $n$  dimensional vector time series  $\{x_t\}_{t=1}^{\infty}$  be cointegrated of order (1,1) with  $r < n$  cointegration vectors. It is assumed that  $x_t$  is generated according to a vector autoregression of order  $p$ , i.e., of the form in (4.9). Let the columns of the  $n \times r$  matrix  $\alpha$  contain the cointegration vectors. Furthermore, let  $M = [S'_k \ \alpha]'$  be an  $n \times n$  nonsingular matrix, where  $S_k = [I_k \ 0]$  is  $k \times n$  and  $k = n - r$  (cf. Theorem 2.1). Finally, let  $A_k^{(c)}$  be determined as in (4.11), where (c) denotes cointegration based parameters, except that  $M$  is as stated here and  $D_{\perp}$  is  $n \times n$  and given by

$$D_{\perp} = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence,  $A^{(c)}(1) = I_n - \sum_{k=1}^p A_k^{(c)} = \gamma\alpha'$  has rank  $r$  under the null hypothesis (see, e.g., Engle and Granger [26] and Johansen [44]) and  $\gamma$  is an  $n \times r$  matrix with  $nr$  free parameters.

In relation to the term structure model, if interest rates are cointegrated of order (1,1) with one cointegration vector such that the spread is stationary, then  $\alpha' = [-1 \ 1]$  while the  $2 \times 1$  vector  $\gamma$  is free. Hence,  $n = 2$ ,  $r = 1$ , and  $k = 1$ . Furthermore,  $S_1 = [1 \ 0]$  is admissible since it has rank 1 and its row is not a cointegration vector.

The null hypothesis I first wish to consider can be stated as follows: The cointegration vectors are fully specified by the matrix  $\alpha$ . Letting  $\Sigma^{(c)}$  denote the matrix  $\Sigma$  conditional on the null,  $\text{vec}$  the column stacking operator,  $\text{vech}$  the corresponding operator which only stacks elements on and below the diagonal of a square matrix (cf. Chapter 2),  $\xrightarrow{d}$  denotes convergence in distribution,  $T$  the sample size of the estimated parameters,  $\mathcal{N}$  the normal distribution, and a hat estimates which are not conditioned on any economic theory restrictions,<sup>11</sup> we can now state the following:

**Proposition 4.4** *Suppose  $\{x_t\}_{t=1}^{\infty}$  is cointegrated of order (1,1) with  $r < n$  cointegration vectors which are fully specified by the matrix  $\alpha$ ,*

$$T^{1/2} \begin{pmatrix} \text{vec}[\hat{A}_1 \cdots \hat{A}_p] - \text{vec}[A_1 \cdots A_p] \\ \text{vech}[\hat{\Sigma}] - \text{vech}[\Sigma] \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, V),$$

where  $A_p D_{\perp} = A_p$ , and suitable regularity conditions are satisfied by the loglikelihood function of the i.i.d. Gaussian sequence  $\{\varepsilon_t\}_{t=1}^T$ , then

$$\text{LR}_h := T \ln \left( \frac{\det[\hat{\Sigma}^{(c)}]}{\det[\hat{\Sigma}]} \right) \xrightarrow{d} \chi_{n(n-r)}^2. \quad (4.24)$$

Instead of proving this proposition formally, I shall present a simple argument why it is true.<sup>12</sup> If  $\alpha$  is known, then under the null hypothesis a Gaussian maximum likelihood

<sup>11</sup>To avoid ambiguities, I do not refer to cointegration restrictions as being economic theory restrictions.

<sup>12</sup>Suitable regularity conditions on the Gaussian loglikelihood function are given by, e.g., Amemiya [3, pages 142–143] and Rao [75, pages 415–420].

estimator of  $(\{B_k\}_{k=1}^p, \Omega)$  has a joint nondegenerate asymptotic normal distribution (see, e.g., Baillie [4]). The reason for this is, of course, that the parameters are estimated from jointly wide sense stationary time series. Accordingly, if we compute the  $A_k$  matrices from the  $B_k$  matrices as in (4.11) and note that  $\Sigma^{(c)} = M^{-1}\Omega(M^{-1})'$ , the constrained maximum likelihood estimator of  $(\{A_k\}_{k=1}^p, \Sigma)$  has a limiting normal distribution. In addition, if the unconstrained estimator of  $(\{A_k\}_{k=1}^p, \Sigma)$  has a limiting normal distribution, we are clearly in the standard setting for performing classical inference. Hence, we can expect, e.g., the likelihood ratio test to have a limiting  $\chi^2$  distribution. It should be emphasized, however, that such an unconstrained estimator often does not exist under the null hypothesis (cf. Sims, Stock and Watson [92]), thus limiting the possible applications of the proposition. Nevertheless, it can be made use of here as well as in any bivariate vector autoregressive setting where both variables are associated with nonzero cointegration coefficients.<sup>13</sup>

The number of degrees of freedom of the test statistic in (4.24) can be calculated as follows. If the  $n \times r$  matrix  $\alpha$  is known, then  $A^{(c)}(1) = \gamma\alpha'$  has  $nr$  free parameters, i.e., the number of (free) parameters of  $\gamma$ . In the alternative model, the matrix  $A(1)$  can be of full rank. Accordingly, it has  $n^2$  free parameters. Hence, the number of degrees of freedom of the  $LR_h$  test is  $n^2 - nr = n(n - r)$ .<sup>14</sup>

For the data generating model of the yield series, letting  $\alpha' = [-1 \ 1]$  under the null hypothesis, we know from the discussion in section 4.3.1 that we can estimate the parameters in (4.9) without conditioning on cointegration and the (least squares or maximum likelihood) estimator has a joint nondegenerate asymptotic normal distribution. Hence, the conditions of Proposition 4.4 are satisfied and the LR test in (4.24) has a limiting  $\chi^2$  distribution with 2 degrees of freedom.

It may be noted that Wald and LM tests may also be considered in the present setting. However, the LR test seems far easier to compute. More importantly, the approach I consider for testing cointegration allows us to derive a *joint* test of the term structure and the cointegration restrictions which is very easy to apply in practise. To my knowledge, no other cointegration test is based on such a joint hypothesis.<sup>15</sup> In particular, letting a bar represent an estimator which is conditioned on some economic theory restrictions, the following holds true:

---

<sup>13</sup>A rule of thumb *appears* to be that if cointegration is limited to be of order (1,1) and  $\alpha$  can be written such that all rows are associated with at least one nonzero coefficient, then an unconstrained maximum likelihood estimator of  $(\{A_k\}_{k=1}^p, \Sigma)$  has a joint nondegenerate asymptotic normal distribution. Whether this is true or not remains (to my knowledge) to be shown.

<sup>14</sup>The general idea underlying Proposition 4.4 is examined by Johansen and Juselius [49, Theorem 2]. They show that we can test a partially or a fully specified  $\alpha$  matrix with an LR test under weaker conditions. In particular, the assumption of an asymptotically normal unconstrained estimator of the  $A_k$  matrices can be dispensed with. In their model, if  $\alpha$  is fully specified, then the LR test has a limiting  $\chi^2$  distribution with  $r(n - r)$  degrees of freedom, where the lower number of degrees of freedom for their statistic depends on the fact that their "alternative" model is based on  $r$  cointegration vectors, whereas the alternative model here is based on  $n$  cointegration vectors. The primary reasons why I consider Proposition 4.4 are: (i) the test statistic in (4.24) is substantially easier to compute than the test statistic in Johansen and Juselius, and (ii) it naturally introduces the main test statistic below (cf. equation (4.26)).

<sup>15</sup>The Johansen procedure only allows for tests on the cointegration vectors (so far, at least). It seems likely, however, that his approach can also deal with tests on the short run dynamics.

**Proposition 4.5** *Suppose that in addition to the assumptions in Proposition 4.4 the  $q \times (np - r)$  affine matrix constraint function  $F(A_1, \dots, A_p) = 0$  and the  $q \times n(p - 1)$  affine matrix constraints  $G(B_1, \dots, B_{p-1}) = 0$  are consistent with the restrictions in (4.11), and*

$$T^{1/2} \begin{pmatrix} \text{vec}[\bar{A}_1 \cdots \bar{A}_p] - \text{vec}[A_1 \cdots A_p] \\ \text{vech}[\bar{\Sigma}] - \text{vech}[\Sigma] \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \bar{V}),$$

then

$$\text{LR}_b := T \ln \left( \frac{\det[\bar{\Sigma}^{(c)}]}{\det[\bar{\Sigma}]} \right) \xrightarrow{d} \chi_{(n-q)(n-r)}^2, \quad (4.25)$$

and

$$\text{LR}_{bh} := T \ln \left( \frac{\det[\bar{\Sigma}^{(c)}]}{\det[\bar{\Sigma}]} \right) \xrightarrow{d} \chi_{(qn[p-1]+n(n-r))}^2. \quad (4.26)$$

Note that the test statistic in (4.25) is conditioned on the economic theory restrictions  $F(A_1, \dots, A_p) = 0$  under the null and the alternative, whereas  $G(B_1, \dots, B_{p-1}) = 0$  are only consistent with the null, i.e., the latter constraints are preconditioned on the assumed form of nonstationarity (cf. equations (4.19)–(4.21)). A rejection of the null hypothesis is therefore due to the cointegration restrictions (assuming the data generating model is not misspecified). The number of degrees of freedom are obtained from noting that, under the alternative,  $(n - q)n$  parameters of  $A(1)$  are free, while  $(n - q)r$  are free under the null. The number of degrees of freedom are thus  $(n - q)(n - r)$ . If  $q = 0$  this corresponds to what we obtained for  $\text{LR}_h$  in (4.24).

The test statistic in (4.26), on the other hand, is only conditioned on the economic theory restrictions under the null hypothesis. Hence, a rejection of the null implies that either the economic theory or the cointegration restrictions are inconsistent with data. It should be noted that the proposition specifies  $p - 1$  lags in the  $y_t$ -model. This implies that the  $x_t$ -model has  $p$  lags and  $A_p D_{\perp} = A_p$ . The alternative model in the proposition has  $n^2 p - nr$  free parameters, while the null model has  $(n - q)(np - r) - (n - q)(n - r)$  free parameters. The latter follows from noting that the economic theory restrictions  $F(\cdot) = 0$  leave the  $A_k$  matrices with  $(n - q)n$  free parameters each, except  $A_p$  which only has  $(n - q)(n - r)$  free parameters. Furthermore, the economic theory restrictions imply that there are  $(n - q)n$  free parameters of  $A(1)$ . Since these restrictions are consistent with the form of cointegration under the null, it follows that cointegration and the economic theory restrictions provide  $A(1)$  with  $(n - q)r$  free parameters. The number of degrees of freedom is thus  $qn(p - 1) + n(n - r)$ .<sup>16</sup>

Finally, if an estimator of  $(\{A_k\}_{k=1}^p, \Sigma)$  which is not constrained on any restrictions converges to a joint nondegenerate asymptotic normal distribution, an estimator which is conditioned on the  $F(\cdot) = 0$  restrictions but not on  $\alpha$  generally also has a limiting normal distribution (see, e.g., Baillie [4] and Warne [99]). Hence, the additional assumptions in

<sup>16</sup>On the other hand, if we let all columns of  $A_p$  be nonzero under the alternative, we have that  $B_p D_{\perp} = 0$  must hold, i.e., only the final  $r$  columns of  $B_p$  may be nonzero. In that case, the alternative model has  $n^2 p$  free parameters while the null model has  $(n - q)np - (n - q)(n - r)$  parameters which are free. Accordingly, the number of degrees of freedom of the  $\text{LR}_{bh}$  test is then  $qnp + (n - q)(n - r)$ , i.e., there are  $qr$  extra restrictions. The reason is simply that we have  $nr$  more free parameters in the alternative model, while the economic restrictions imply that only  $(n - q)r$  of these additional parameters are free.

Proposition 4.5 do not seem strong and we expect  $LR_b$  and  $LR_{bh}$  to have a limiting  $\chi^2$  distribution whenever  $LR_h$  does.<sup>17</sup>

In relation to the term structure model, the affine matrix constraint functions  $F(\cdot) = 0$  and  $G(\cdot) = 0$  correspond to equations (4.17) and (4.18), respectively, if we move the matrices on the right hand side to the left hand side. Since  $n = 2$  and  $q = 1$ , the  $LR_{bh}$  test has  $2p$  degrees of freedom for the joint test of cointegration and the term structure model when equation (4.10) is associated with  $p - 1$  lags.

## 4.4 Empirical Results

### 4.4.1 The Data

The observations on the two yield series I shall investigate are depicted in Figure 4.1. These are monthly rates on 1-month Treasury bills (Statsskuldväxlar) and 5-year Treasury bonds (Riksobligationer) from Swedish bond markets, where the latter series represents the consol yield. The sample covers the period November 1983 to December 1989.<sup>18</sup> In order to obtain the best possible data on the 1-month yield, I have chosen quotations with delivery on the 15:th whenever possible. Since delivery is always two banking days after trade in the bill has taken place, the day when the actual observation is made is generally on the 13:th.<sup>19</sup> The observations on 5-year bond yields are then taken from the same day as those for the 1-month bill.

5-year Treasury bonds were introduced in November 1983. Since trade in other long term bonds was very limited prior to that month, reliable observations on a long term yield do not exist for the period up to November 83. Furthermore, the trading volume in bonds with a time to maturity greater than 5 years has been low relative to that of 5-years bonds for the sample period. Hence, this yield series is the most reliable source on a long term rate.

### 4.4.2 Tests of the Term Structure Model in Swedish Bond Markets

Before we select a value for the lag order parameter,  $p$ , and test the restrictions in (4.17) and (4.18), it should be noted that I have adjusted the interest rate series around estimates of their unconditional means. The primary reason for doing this is simply that I wish to restrict the empirical analyses such that linear deterministic trends are not allowed for. If interest rates "are" jointly wide sense stationary, this is not a restriction on the estimated

---

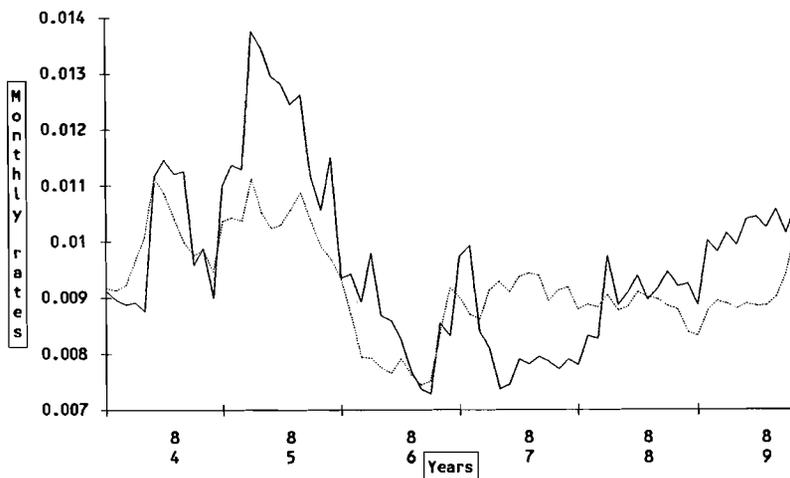
<sup>17</sup>Under the assumptions considered by White [103], the Wald and the LM statistics have an asymptotic  $\chi^2$  distribution even when the econometric model is misspecified. This, however, is not true for the LR test.

<sup>18</sup>The data have been collected by Svenska Handelsbanken. I am grateful to Greger Wahlstedt for making data easily accessible to Ossian and myself.

<sup>19</sup>Sometimes the 15:th is on a non banking day. In these cases I have looked at the closest banking day to determine the delivery date. The reason why I consider this dating scheme is simply that Treasury bills are generally issued, as 6-month or 12-month bills, in the middle of each month (between the 6:th and the 20:th). Thus, end of month observations are associated with anything from 1-week to 7-week bills, whereas middle of month observations are associated with 3-week to 5-week bills.

parameters. On the other hand, if interest rates “are” cointegrated of order (1,1) with a stationary spread, the constant vector  $\mu$  in (4.9) is related to linear deterministic trends in  $x_t$  unless  $\mu = \gamma\mu_0$ , where  $\gamma$  is a  $2 \times 1$  vector determined as in section 4.3.4 (cf. Johansen [44]). A simple procedure for imposing such restrictions is to measure  $x_t$  about an estimate of its unconditional mean (see also the appendix).<sup>20</sup>

Figure 4.1: Monthly yield to maturity on a 1-month Treasury bill (continuous line) and a 5-year Treasury bond (dotted line) in Sweden, 1984:2–1989:12.



The procedure I consider for determining a suitable lag order in (4.9) and (4.10) is based on examining three information criteria and two lag order tests. In Paulsen [71] it is shown that the log criterion (SIC) of Schwarz [84] and the iterated log criterion (ILC) of Hannan and Quinn [39] are consistent measures for determining the true lag order even in the presence of unit roots. The Akaike information criterion (AIC) (cf. Akaike [2]), however, is not consistent. For finite samples, it is well known that AIC tends to overestimate the true lag order, whereas SIC and to some extent ILC tends to underestimate it (see, e.g., Hannan and Quinn [39] and Nickelsburg [68]). Accordingly, I let these three criteria pick an upper and a lower bound for  $p$  and expect SIC to give me the lower and AIC the upper. Within this set of permissible lag orders, I test hypotheses of the form  $p = s$  versus  $p > s$  (in Table 4.1 this is referred to as  $s$  versus  $p$  lags) using the Wald and the LR tests. In the present setting, these tests have, as shown by Sims, Stock and Watson [92, page 134] (see also the appendix), a limiting  $\chi^2$  distribution.

<sup>20</sup>In fact, the constant  $\mu_0$  is equal to the term premium if the term structure model is consistent with the cointegrated vector autoregression in equation (4.10).

Table 4.1: Lag order selection and Granger–noncausality tests for data generating processes of  $x_t$  and  $y_t$  in Sweden, 1983:11+ $p$ –1989:12.

Information Criteria:

$p$	$x_t$			$y_t$		
	SIC	ILC	AIC	SIC	ILC	AIC
1	-30.181*	-30.492*	-30.651	-30.126*	-30.440*	-30.601
2	-29.860	-30.487	-30.810	-29.763	-30.397	-30.724
3	-29.394	-30.345	-30.835	-29.342	-30.303	-30.799
4	-28.950	-30.231	-30.892	-28.870	-30.165	-30.834
5	-28.488	-30.106	-30.942	-28.396	-30.031	-30.878
6	-28.015	-29.977	-30.994	-28.005	-29.988	-31.017
7	-27.689	-30.003	-31.203	-27.654	-29.993	-31.209
8	-27.339	-30.013	-31.402*	-27.318	-30.021	-31.428*
9	-27.057	-30.098	-30.681	-26.816	-29.892	-30.495
10	-26.522	-29.939	-30.720	-26.244	-29.700	-30.505

Lag Order Tests:

	1 vs. 2	1 vs. 3	1 vs. 4	2 vs. 3	2 vs. 4	3 vs. 4	
$x_t$	LR	11.568	14.902	19.466	3.888	8.482	4.732
	(p-value)	(.021)	(.061)	(.078)	(.421)	(.388)	(.316)
	W	12.329	15.888	20.942	3.976	8.845	4.885
(p-value)	(.015)	(.044)	(.051)	(.409)	(.356)	(.299)	
$y_t$	LR	10.299	15.908	19.924	6.098	10.286	4.369
	(p-value)	(.036)	(.044)	(.068)	(.192)	(.245)	(.358)
	W	10.791	16.946	21.468	6.312	10.837	4.501
(p-value)	(.029)	(.031)	(.044)	(.177)	(.211)	(.342)	

Granger–Noncausality Tests:

		$p = 1$	$p = 2$	$p = 3$	$p = 4$
$r_t$ -equation	LR	4.057	6.045	6.046	9.044
	(p-value)	(.044)	(.049)	(.109)	(.060)
	W	4.173	6.310	6.315	9.663
(p-value)	(.041)	(.043)	(.097)	(.046)	
$R_t$ -equation	LR	1.435	5.898	6.443	7.455
	(p-value)	(.231)	(.052)	(.092)	(.114)
	W	1.450	6.150	6.749	7.872
(p-value)	(.228)	(.046)	(.080)	(.096)	
$\Delta r_t$ -equation	LR	7.290	10.045	11.790	12.146
	(p-value)	(.007)	(.006)	(.008)	(.016)
	W	7.672	10.791	12.841	13.281
(p-value)	(.006)	(.004)	(.005)	(.010)	
$(R_t - r_t)$ -equation	LR	.869	2.627	6.306	10.010
	(p-value)	(.351)	(.269)	(.098)	(.040)
	W	.874	2.676	6.599	10.773
(p-value)	(.350)	(.262)	(.086)	(.029)	

Note: The \* indicates the minimum value of an information criterion for a maximum lag order of 15. The p-value for a lag order test of the hypothesis  $s$  versus  $p$  lags is taken from the  $\chi^2$  distribution with  $4(p-s)$  degrees of freedom. A Granger–noncausality test in, e.g., the  $r_t$ -equation is a test of the hypothesis:  $\alpha_{12k} = 0$  for all  $k \in \{1, \dots, p\}$ . The p-value for such a test is taken from a  $\chi^2$  distribution with  $p$  degrees of freedom.

From Table 4.1 we find that SIC and ILC are both minimized at  $p = 1$  and the Akaike criterion at  $p = 8$  for the  $x_t$  and the  $y_t$  models. It can be seen that for each information statistic the actual value, given  $p$ , is generally smaller in the  $x_t$ -model. Furthermore, whereas the differences between the two smallest values for SIC and AIC are quite large, this is not true for ILC. For both models,<sup>21</sup> the second smallest value of ILC is obtained at  $p = 2$  and particularly the  $x_t$ -model is associated with a small difference between the value of ILC at lag orders 1 and 2. As we shall see below, the choice between these two lag orders leads to sharply different conclusions on the term structure tests.

In the box below the information criteria of Table 4.1 a sample of lag order tests is given. Here, the cautious attitude towards selecting  $p$  solely from an information criterion is to some extent confirmed. For example, in VAR(2) models for  $x_t$  and  $y_t$ , the null hypotheses  $A_2 = 0$  and  $B_2 = 0$  do not receive much support from the data. Furthermore, when testing a VAR(2) model within VAR(3) and VAR(4) models, the test statistics are not significantly different from zero. This suggests that  $p = 2$  should be chosen.

The bottom box of Table 4.1 contains results from various tests of Granger–noncausality. In the  $x_t$ -model we find some evidence that the long term yield may Granger–cause the short term yield. Also, in a VAR(2)-model both test statistics suggest that the short term rate may Granger–cause the long term rate. Since the estimate of  $\rho$  is .991 we know from the discussion in section 4.3.3 that the null hypothesis  $a_{21k} = 0$  for all  $k \in \{1, \dots, p\}$  may be difficult to reject when data are consistent with the term structure model. Furthermore, in view of Proposition 4.2 these results are quite encouraging for testing the term structure restrictions in a VAR(2) model. Also, the Granger–noncausality tests in the  $r_t$ -equation of a VAR(3) model suggest that a failure to reject the term structure model may either be associated with large uncertainty of the estimated coefficients or with explosive roots.

If we turn to the  $y_t$ -model, it can be shown from equation (4.11) that  $b_{12k} = a_{12k}$  for all  $k$  when interest rates are cointegrated (see also (4.21)). On the other hand,  $a_{211} = (1 + b_{111} - b_{121} + b_{211} - b_{221})$  and  $a_{21k} = (b_{11,k} - b_{11,k-1} - b_{12,k} + b_{21,k} - b_{21,k-1} - b_{22,k})$  for all  $k \in \{2, \dots, p\}$ , i.e., the hypothesis  $b_{21k} = 0$  for all  $k$  is not equivalent to the hypothesis  $a_{21k} = 0$  for all  $k$ . Yet, the Granger–noncausality tests in the  $(R_t - r_t)$ -equation are of some interest. From equations (4.20) and (4.21) and the assumption that  $\{x_t\}_{t=1}^{\infty}$  is not an explosive stochastic process, it can be shown that if  $\{\Delta r_{t-s}\}_{s=1}^p$  does not Granger–cause  $(R_t - r_t)$  and the term structure model is consistent with data, then  $(1 - \rho)/\rho \leq b_{121} \leq (1 + \rho)/\rho$ . Hence, the hypothesis  $b_{21k} = 0$  for all  $k$  and nonexplosive interest rates provides us with a set of permissible positive values for  $b_{121}$  and, thus, for  $a_{121}$  when data are consistent with the term structure model. In Table 4.1 it can be seen that for  $p \leq 3$ , the hypothesis of Granger–noncausality in the  $(R_t - r_t)$ -equation is not rejected while the hypothesis is strongly rejected in the  $\Delta r_t$ -equation.

To summarize the analyses so far, the lag order selection scheme suggests that  $p = 2$  should be chosen for  $x_t$  and  $y_t$ . Second, the Granger–noncausality tests provide evidence which suggests that tests of the term structure restrictions will not suffer from substantial uncertainty about the estimated parameters, i.e., the tests indicate that the long term rate Granger–causes the short term rate and, for a VAR(2) model, that the opposite may also be consistent with data.

---

<sup>21</sup>It is perhaps misleading to view  $x_t$  and  $y_t$  as merely two models since different lag orders are associated with different vector autoregressive models. Nevertheless, the “sloppy” terminology is used.

Table 4.2: Tests of the term structure restrictions and estimated coefficients for VAR(2) data generating processes of  $x_t$  and  $y_t$  in Sweden, 1984:2–1989:12.

\*\*\* The  $x_t$  Data Generating Model \*\*\*

$$\begin{aligned} E[r_t] &= .961 \cdot 10^{-2} & E[R_t] &= .926 \cdot 10^{-2} \\ V[r_t] &= .225 \cdot 10^{-5} & V[R_t] &= .737 \cdot 10^{-6} \\ R^2 &= .765 & R^2 &= .845 \end{aligned}$$

$\hat{\rho} = .99083$

$k$	$\hat{a}_{11k}$	$\hat{a}_{12k}$	$\hat{a}_{21k}$	$\hat{a}_{22k}$	$\bar{a}_{11k}$	$\bar{a}_{12k}$	$\bar{a}_{21k}$	$\bar{a}_{22k}$
1	.549 (.140)	.643 (.296)	-.156 (.065)	1.364 (.137)	.719 (.118)	.233 (.249)	-.009 (-)	1.009 (-)
2	.250 (.139)	-.393 (.302)	.149 (.065)	-.453 (.141)	.077 (.117)	.139 (.255)	.0 (-)	.0 (-)

$$\hat{\Sigma} = \begin{bmatrix} .528 & (.089) & .132 & (.033) \\ .132 & (.033) & .114 & (.019) \end{bmatrix} \cdot 10^{-6} \quad \bar{\Sigma} = \begin{bmatrix} .560 & (.094) & .160 & (.038) \\ .160 & (.038) & .138 & (.023) \end{bmatrix} \cdot 10^{-6}$$

convergence measure for constrained estimates after 3 iterations:  $.177 \cdot 10^{-31}$

$$\ln L(\hat{A}_1, \hat{A}_2, \hat{\Sigma}) = 1092.794$$

$$\ln L(\bar{A}_1, \bar{A}_2, \bar{\Sigma}) = 1085.936$$

Tests of Term Structure Restrictions:

$$\begin{aligned} LR(4) &= 13.717 & \text{p-value} &= .008 \\ W(4) &= 15.131 & \text{p-value} &= .004 \\ W^*(4) &= 12.198 & \text{p-value} &= .016 \end{aligned}$$

\*\*\* The  $y_t$  Data Generating Model \*\*\*

$$\begin{aligned} E[\Delta r_t] &= .007 \cdot 10^{-3} & E[(R_t - r_t)] &= -.366 \cdot 10^{-3} \\ V[\Delta r_t] &= .611 \cdot 10^{-8} & V[(R_t - r_t)] &= .177 \cdot 10^{-5} \\ R^2 &= .152 & R^2 &= .678 \end{aligned}$$

$\hat{\rho} = .99083$

$k$	$\hat{b}_{11k}$	$\hat{b}_{12k}$	$\hat{b}_{21k}$	$\hat{b}_{22k}$	$\bar{b}_{11k}$	$\bar{b}_{12k}$	$\bar{b}_{21k}$	$\bar{b}_{22k}$
1	.188 (.241)	.608 (.283)	.063 (.206)	.823 (.242)	-.083 (.206)	.153 (.242)	.083 (.206)	.856 (.242)
2	.142 (.111)	-.401 (.296)	-.151 (.095)	-.008 (.253)	.152 (.095)	.041 (.252)	-.152 (.095)	-.041 (.252)

$$\hat{\Omega} = \begin{bmatrix} .518 & (.087) & -.388 & (.070) \\ -.388 & (.070) & .378 & (.063) \end{bmatrix} \cdot 10^{-6} \quad \bar{\Omega} = \begin{bmatrix} .539 & (.090) & -.389 & (.071) \\ -.389 & (.071) & .378 & (.063) \end{bmatrix} \cdot 10^{-6}$$

convergence measure for the constrained estimates after 4 iterations:  $.470 \cdot 10^{-33}$

$$\ln L(\hat{B}_1, \hat{B}_2, \hat{\Omega}) = 1090.689$$

$$\ln L(\bar{B}_1, \bar{B}_2, \bar{\Omega}) = 1085.665$$

Tests of Term Structure Restrictions:

$$\begin{aligned} LR(4) &= 10.047 & \text{p-value} &= .040 \\ W(4) &= 10.792 & \text{p-value} &= .029 \\ W^*(4) &= 11.082 & \text{p-value} &= .026 \end{aligned}$$

**Note:** Estimated parameters with a bar satisfy the term structure restrictions whereas those with a hat need not. Estimated standard errors of estimated coefficients are reported within parenthesis.

In Table 4.2 I report estimated coefficients and tests of the term structure restrictions in the  $x_t$  and the  $y_t$  environments. The estimates of the parameters of the data generating processes in (4.9) and (4.10) are obtained from unconstrained and constrained maximum likelihood (see, e.g., Baillie [4], Berndt and Savin [9], Oberhofer and Kmenta [69], Rothenberg [79], and Warne [99]). In addition to the LR and Wald tests, the  $W^*$ -test is a Wald test where White's [102, pages 133–135] heteroskedasticity consistent covariance matrix estimator is used. For the  $x_t$ -model we find very little or no evidence that the term structure restrictions are supported by data. Looking at individual coefficients, it can be seen that the constraints change the value of  $a_{222}$  to a value outside a 95 percent confidence bound around  $\hat{a}_{222}$ , while all other constrained estimates are within similar confidence bounds around the unconstrained estimates. Similarly, if we assume that  $\varepsilon_t$  is i.i.d. Gaussian we can easily calculate standard errors for each element of an estimate of  $\Sigma$  and  $\Omega$  (cf. Magnus and Neudecker [63, Theorem 15.4]). In Table 4.2 these standard errors are reported within parenthesis in the estimated  $\Sigma$  and  $\Omega$  matrices. For  $\bar{\Sigma}$  we have that each individual element is well within a 95 percent confidence bound about the corresponding element of  $\hat{\Sigma}$ .

For the  $y_t$ -model we find that none of the constrained estimates of  $b_{ijk}$  are outside 95 percent confidence bounds about the unconstrained estimates. In view of this, it is perhaps not surprising to find that the p-values of the term structure tests are much larger for the  $y_t$ -model than for the  $x_t$ -model. Hence, if interest rates are assumed to be cointegrated of order (1,1) with a stationary spread the evidence against the term structure model as given by the test statistics is not entirely convincing. Given the results from the  $x_t$ -model, the test results thus far are somewhat ambiguous. Let us therefore look at some graphs on theoretical and observed long term rate series (Figure 4.2).

The theoretical yield series have been computed in the manner proposed by Campbell and Shiller [14, equation (9)]. That is,

$$\begin{aligned}\tilde{R}_t &:= (1 - \hat{\rho})N_1J_2(I_4 - \hat{\rho}\hat{A})^{-1}X_t, \\ \tilde{R}_t - \tilde{r}_t &:= \hat{\rho}N_1J_2\hat{B}(I_4 - \hat{\rho}\hat{B})^{-1}Y_t,\end{aligned}$$

$Y_t = [y'_t \ y'_{t-1}]'$  and  $B$  has the same form as  $A$  except that  $A_k$  has been replaced by  $B_k$ .

From Figure 4.2 we find that the  $\tilde{R}_t$  series tracks the  $R_t$  series very poorly, whereas the  $(\tilde{R}_t - \tilde{r}_t)$  series follows  $(R_t - r_t)$  quite closely. In relation to Campbell and Shiller's study, this latter observation is consistent with what they have found on U.S. data. Keeping the above test results in mind, it may seem as if this is precisely what we should expect. However, the great difference between the two graphs in Figure 4.2 may also be a result of some numerical problem. To examine whether the cointegration restrictions are behind this I have computed an  $\tilde{R}_t$  series based on  $\hat{B}_1$  and  $\hat{B}_2$ , using (4.11). However, that gives us a series which closely resembles the one found in the upper chart of Figure 4.2. Accordingly, a plausible reason seems to be that the constant  $(1 - \hat{\rho}) \approx 0$ , which directly appears in the  $\tilde{R}_t$  but *not* in the  $(\tilde{R}_t - \tilde{r}_t)$  expression, is responsible for the difference. Hence, the variability of the theoretical series on the long term yield is probably very sensitive to which one of these expression we apply, thereby lowering the usefulness of such expressions for making economic interpretations. In particular, to conclude that the bottom graph in Figure 4.2 provides some encouraging evidence for the term structure model seems somewhat overly optimistic.

Figure 4.2: Actual (continuous line) and theoretical (dotted line) observations on the long term yield (top chart) and the spread between the long term and short term yield (bottom chart) in Sweden, 1984:2-1989:12.

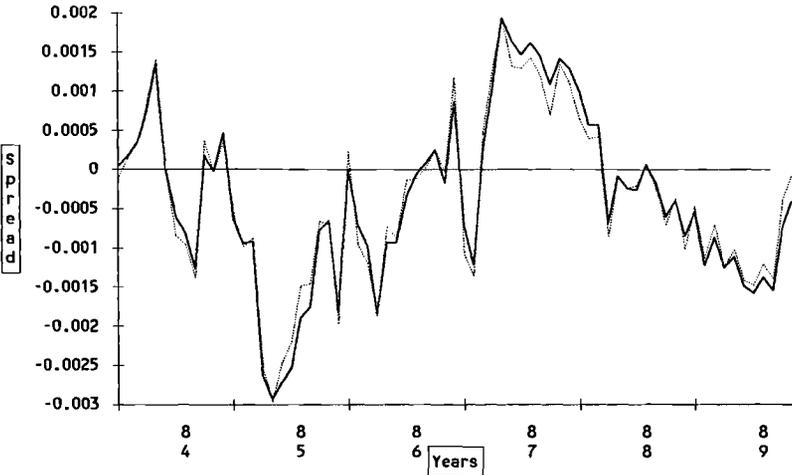
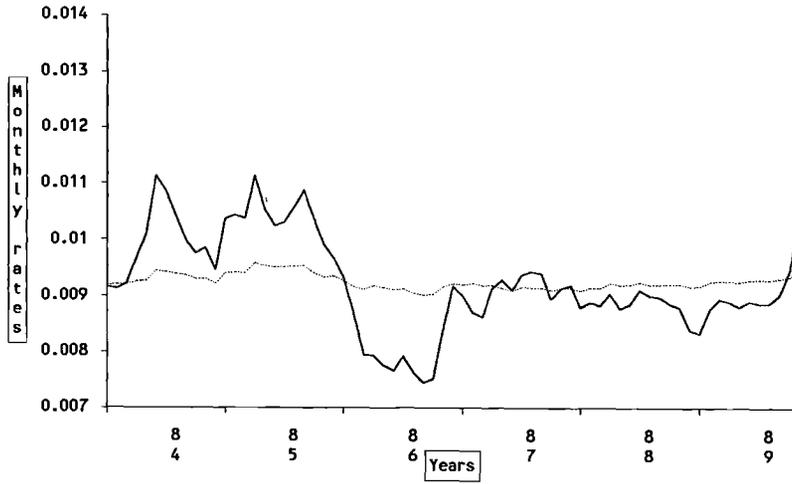


Table 4.3: Likelihood ratio tests of cointegration and the term structure restrictions when data on  $x_t$  and  $y_t$  are generated by a VAR(2) model, 1984:2–1989:12.

Hypothesis:	$\hat{\Sigma}^{(c)}$ vs. $\bar{\Sigma}$	$\bar{\Sigma}^{(c)}$ vs. $\bar{\Sigma}$	$\hat{\Sigma}^{(c)}$ vs. $\bar{\Sigma}$
LR	5.925	2.215	15.971
df	2 ( <i>h</i> )	1 ( <i>b</i> )	6 ( <i>bh</i> )
p-value	.052	.137	.014

**Note:** The hat denotes estimates which are not constrained on the term structure model while the bar corresponds to estimates which are based on these constraints. The superscript (*c*) relates to estimates which are conditioned on the cointegration restrictions, i.e., one cointegration vector such that  $\{(R_t - r_t)\}_{t=1}^{\infty}$  is stationary. The degrees of freedom (df) for the limiting  $\chi^2$  distribution are obtained from Propositions 4.4 and 4.5.

So far I have discussed the term structure tests as if they are, in some sense, equivalent for the  $x_t$  and the  $y_t$  models. However, whereas a test in the former model is not conditioned on a certain kind of nonstationarity, a test in the latter model is. In Table 4.3 we find tests based on Propositions 4.4 and 4.5. Two of these tests only examine the cointegration restrictions while the third statistic is employed to test the joint hypothesis of cointegration and the term structure model. If we condition on the term structure model when testing the cointegration restrictions, we find that the  $LR_b$  test has a p-value around 14 percent, whereas the  $LR_h$  test, which is not conditioned on the term structure restrictions, has a p-value around 5 percent. Hence, the cointegration restrictions in isolation are only weakly supported by data. Moreover, when we test the term structure model and the cointegration constraints jointly, the  $LR_{bh}$  test has a p-value around 1.4 percent. This is less than half of what was found for the tests of the term structure model conditional on cointegration. It should be noted that the value of the LR test is, for obvious reasons, higher in Table 4.3 than in Table 4.2. Yet, there is no a priori reason to expect the p-value to be lower in the former case since this test is associated with a larger number of restrictions.

To sum up, from the analyses of the term structure model (when data are generated by a VAR(2) model) one may conclude that it generally receives very little support from Swedish interest rate data. However, if the empirical analyses is based on a prior “belief” in cointegrated interest rates, the term structure tests provide some support for the model, except when the null hypothesis also incorporates the cointegration restrictions. Since economic theory seldom gives any guidance on whether data should be modelled as cointegrated or not, it is generally preferable to test the cointegration restrictions along with the term structure restrictions. I shall return to these issues in the next section.

Before coming to that, in Table 4.4 I report tests of the term structure restrictions for various choices of lag order. In the  $x_t$ -model we find that the conclusion from these tests is particularly sensitive to the decision between  $p = 1$  and  $p \geq 2$ . For cointegrated interest rates, the LR and, to some extent, the Wald test appear to be sensitive to the choice between  $p = 2$  versus lag orders in the immediate vicinity. In view of these results, we may conclude that the choice of  $p$ , which is *not* a trivial matter, can be very important for the conclusion from the term structure tests. Furthermore, for the heteroskedasticity consistent Wald test the conclusion changes dramatically for both models when we examine a lag order greater than one in relation to  $p = 1$ . Finally, Campbell and Shiller [14, Table 3] report a p-value of .03 percent (the full sample on U.S. data) for a  $W^*$ -test of the cross equation restrictions in (4.18) when the Akaike criterion is employed for selecting  $p$ .

Table 4.4: Tests of the term structure restrictions for VAR( $p$ ) data generating processes of  $x_t$  and  $y_t$  in Sweden, 1984:p–1989:12.

$p$	LR( $2p$ )		W( $2p$ )		W*( $2p$ )	
	$x_t$	$y_t$	$x_t$	$y_t$	$x_t$	$y_t$
1	3.873 (.144)	2.158 (.341)	3.979 (.137)	2.184 (.336)	4.281 (.118)	2.295 (.317)
2	13.717 (.008)	10.047 (.040)	15.131 (.004)	10.792 (.029)	12.198 (.016)	11.082 (.026)
3	13.853 (.031)	11.247 (.081)	15.319 (.018)	12.201 (.058)	19.666 (.003)	17.123 (.009)
4	14.919 (.061)	11.622 (.169)	16.655 (.034)	12.658 (.124)	20.795 (.008)	21.288 (.006)
8	37.246 (.002)	26.495 (.047)	50.285 (.000)	32.710 (.008)	76.644 (.000)	58.843 (.000)

**Note:** The  $p$ -values reported within parenthesis are taken from the  $\chi^2$  distribution with  $2p$  degrees of freedom.

Had I also made use of such a test procedure, the  $W^*$ -test in the  $y_t$ -model for  $p = 8$  has a  $p$ -value of approximately the same magnitude. If we also take the similarities between Figure 4.2 above and Figure 1 in Campbell and Shiller into account, one may suspect that a less mechanical approach to the selection of  $p$  and a different test statistic can change some of the conclusions made by Campbell and Shiller markedly.

### 4.4.3 Inference from Bootstrap Distributions

So far, all inference has been carried out under the assumption that asymptotic distribution theory provides us with good approximations of the tests' unknown small sample distributions. In general, the limiting  $\chi^2$  distribution of the LM, LR, and Wald tests is derived for a sequence of i.i.d. Gaussian innovations  $\{\varepsilon_t\}_{t=1}^\infty$ . Since it is often argued that innovations to financial variables, such as interest rates and exchange rates, are better modelled as leptokurtic than Gaussian, i.e., the tails of the probability distributions from which the innovations are generated are thicker than those of the Gaussian probability distribution, one may perhaps lean towards questioning the usefulness of a distribution which is based on Gaussian innovations.

It is, of course, possible that leptokurtosis is a reflection of, e.g., conditional heteroskedasticity (see, e.g., Engle, Lilien and Robins [27]). Below, however, I shall limit the analyses from such cases. Instead, I am interested in studying a relevant alternative to the  $\chi^2$  distribution for small samples. The maintained assumption is that  $\{\varepsilon_t\}_{t=1}^\infty$  is i.i.d. The questions I shall focus on in this section are:

- (i) shall *possible* nonstationarity be accounted for prior to testing the economic hypothesis,
- (ii) is the choice of test statistic important, and
- (iii) how does the limiting  $\chi^2$  distribution perform in relation to the alternative estimates of the small sample uncertainty.

To some extent, the first and second question have been addressed above. Here, I shall use bootstrap resampling which, among other things, allows me to study cases when the

estimated model differs from the data generating model. The analyses will be limited to the W and LR tests of the term structure model within a VAR(2) setting.

The data generating models for interest rates are given by  $(\{\bar{\varepsilon}_t\}_{t=1}^{71}, \{\bar{A}_k\}_{k=1}^2, \{x_{-s}\}_{s=0}^1)$  for  $x_t$ , and  $(\{\bar{\eta}_t\}_{t=1}^{71}, \{\bar{B}_k\}_{k=1}^2, \{y_{-s}\}_{s=0}^1)$  for  $y_t$ . The former model is called the  $x_t$ -model and the latter the  $y_t$ -model. The residuals  $\bar{\varepsilon}_t$  and  $\bar{\eta}_t$  are determined from  $\bar{\varepsilon}_t = x_t - \sum_{k=1}^2 \bar{A}_k x_{t-k}$  and  $\bar{\eta}_t = y_t - \sum_{k=1}^2 \bar{B}_k y_{t-k}$ , respectively. The bootstrap algorithms are constructed along the lines suggested by Efron [23, pages 35–36].

Specifically, suppose the true innovations for the sample,  $\{\varepsilon_t\}_{t=1}^{71}$ , are generated from an unknown probability distribution  $f$  on the 2 dimensional Euclidean space, i.e.,

$$\{\varepsilon_t\}_{t=1}^{71} \sim f.$$

For the  $x_t$ -model, an empirical probability distribution  $f_{\text{boot}}$  is given by

$$f_{\text{boot}} : \text{probability mass } 1/71 \text{ on } \bar{\varepsilon}_1, \bar{\varepsilon}_2, \dots, \bar{\varepsilon}_{71}.$$

Second, a Monte Carlo algorithm is used to randomly select 71 integers,  $i_1, i_2, \dots, i_{71}$ , with replacement, from  $\{1, 2, \dots, 71\}$ . Then, a bootstrap series  $\{\tilde{\varepsilon}_t\}_{t=1}^{71}$  is generated by letting  $\tilde{\varepsilon}_t = \bar{\varepsilon}_{i_t}$ ,  $t = 1, 2, \dots, 71$ . Given the initial conditions,  $\{x_{-s}\}_{s=0}^1$ , and the estimated parameters,  $\{\bar{A}_k\}_{k=1}^2$ , I let

$$\tilde{x}_t := \sum_{k=1}^2 \bar{A}_k \tilde{x}_{t-k} + \tilde{\varepsilon}_t,$$

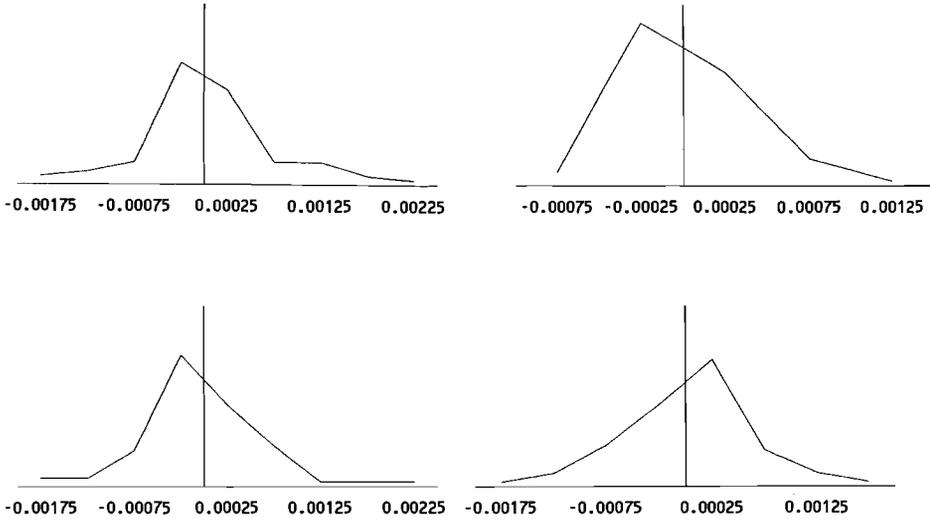
for  $t = 1, 2, \dots, 71$  and  $\tilde{x}_{-s} = x_{-s}$  for  $s = 0, 1$ . In a similar fashion a bootstrap series is constructed within the  $y_t$ -model. These bootstrap series are then transformed to series which correspond to the model which is estimated. For that model, a Wald and an LR test is computed.

The second step is then repeated  $B$  times, yielding ordered sequences  $\{\text{LR}_{(b)}\}_{b=1}^B$  and  $\{\text{W}_{(b)}\}_{b=1}^B$  for each pair of data generating and estimation model. Here, the ordering is such that  $\text{LR}_{(b)} \leq \text{LR}_{(b+1)}$  for every  $b \in \{1, 2, \dots, B-1\}$ , and similarly for the Wald test. To obtain a reasonably good estimate of the unknown small sample distributions for each test statistic, I let  $B = 5000$ . Finally, the *percentile method* is employed to construct the bootstrap distributions (cf. Efron [23, Chapter 10]).<sup>22</sup>

In Figure 4.3 the  $f_{\text{boot}}$  empirical probability distributions for  $\{\bar{\varepsilon}_t\}_{t=1}^{71}$  and  $\{\bar{\eta}_t\}_{t=1}^{71}$  are depicted. The residuals associated with the long term rate equation seem to have thick

<sup>22</sup>The idea behind the percentile method is extremely simple. To construct a 100p percent one-sided confidence bound, I simply let  $p = b/B$ , and for, e.g., the LR test I pick the  $b$ :th largest value. Accordingly, for a given hypothesis, there is no a priori reason why we should expect the bootstrap p-value of the Wald test to be smaller than the bootstrap p-value for the LR test as the case is for the  $\chi^2$  distribution. Furthermore, we do not know how large these p-values are in relation to the limiting  $\chi^2$  p-value. Thus, the bootstrap resampling plan allows us to study questions (ii) and (iii) above. In particular, one of the test statistics can have a severely skewed bootstrap distribution relative to the  $\chi^2$ . If the bootstrap distribution of the LR test is skewed to the right, then  $\chi^2$  based inference for the LR test is less "sensitive" than similar inference for the Wald test and vice versa if the bootstrap distribution of the Wald test is skewed to the left of the  $\chi^2$  distribution. It may also be noted that the percentile method is not the only procedure for constructing bootstrap distributions given, e.g.,  $\{\text{LR}_{(b)}\}_{b=1}^B$  (cf. Efron [23, Chapter 10]). Nevertheless, in many applications it appears to provide quite accurate estimates of the small sample uncertainty.

Figure 4.3: Empirical probability distributions of the constrained residuals for the  $x_t$ -model (top) and the  $y_t$ -model (bottom) with the residuals of the first equation in the graphs to the left and the residuals of the second equation in the graphs to the right. Data is generated by VAR(2) models.

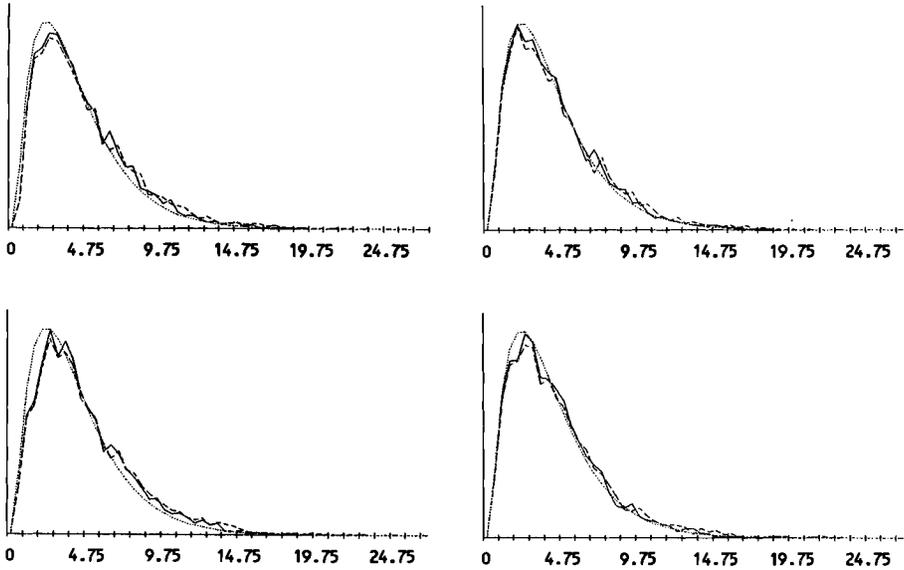


tails, whereas the residuals from the short term rate equation are probably well approximated by a Gaussian distribution. However, as a joint distribution, these two graphs suggest skewness as well as kurtosis. Similarly, the two residual series for the constrained  $y_t$ -model are both skewed, to the left for the  $\Delta r_t$ -equation and to the right for the  $(R_t - r_t)$ -equation, and seem to have one thick tail each. Hence, the joint empirical probability distributions of the  $x_t$  and the  $y_t$  models certainly do not look Gaussian.

The bootstrap distributions of the LR and Wald tests are shown in Figure 4.4 along with the limiting  $\chi^2_4$  distribution. Generally, the bootstrap and  $\chi^2$  distributions appear very similar in all four graphs. Furthermore, since all bootstrap distributions are somewhat skewed to the right, the LR test is less sensitive than the Wald test to the choice of approximation of the unknown small sample distributions. It should perhaps be emphasized that this does not imply that the LR test is better suited for testing the term structure model than the Wald test. However, in view of the fact that the latter test statistic is not numerically invariant in finite samples with respect to nonlinear transformations (such as those associated with the derivation of equations (4.17) and (4.18)) while the LR test is, the distributions in Figure 4.4 suggest that when the  $\chi^2$  distribution is employed for inference, the performance of the LR test seems more reliable.

In Table 4.5 I report p-values for the LR and W tests of the term structure restrictions which are based on three distributions. When we estimate the  $x_t$ -model, the bootstrap

Figure 4.4: Bootstrap distributions for the LR (continuous line) and Wald (dashed line) tests along with the  $\chi_4^2$  (dotted line) distribution. The top charts are associated with the (VAR(2))  $x_t$  data generating model and the charts in the left column with estimated  $x_t$  models.

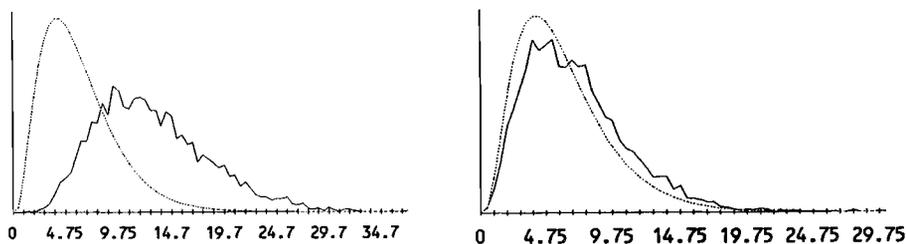


p-values of the  $x_t$  and  $y_t$  data generating models are nearly identical to the  $\chi_4^2$  p-value for the LR test. In contrast, the p-values of the Wald test are identical to those of the LR test when data are generated by the  $x_t$  and the  $y_t$  model, whereas the  $\chi_4^2$  p-value is less than half of the bootstrap p-values. When the  $y_t$ -model is estimated a similar picture emerges. Although all p-values are associated with rejecting the term structure restrictions if a critical 5 percent value is chosen, these results suggest that a cautious attitude regarding the Wald test seems justified. In particular, the size of the Wald test may have been affected by the nonlinear transformations discussed above.

Another interesting observation that can be made from Figure 4.4 is that when the  $x_t$ -model is estimated the bootstrap distributions are somewhat more skewed in relation to the  $\chi^2$  distribution than when the  $y_t$ -model is estimated, but that this skewness seems negligible at the right tails. Furthermore, this does not seem to depend much on whether  $x_t$  or  $y_t$  actually generates data. If the cointegration restrictions in the  $y_t$ -model are nearly “consistent” with the  $x_t$  data generating model, as was discussed for  $\rho \approx 1$  at the end of section 4.3.3, this may explain such behavior of the bootstrap distributions.

To examine this issue further, in Figure 4.5 the bootstrap distributions of the  $LR_{bh}$  tests are depicted when data are generated by the  $x_t$  and  $y_t$  models, along with the density function of the  $\chi_8^2$  distribution. The estimated model for both bootstrap distributions is the  $y_t$ -model. If the cointegration and term structure restrictions are satisfied by the

Figure 4.5: Bootstrap distributions for the  $LR_{bh}$  test (continuous line) and the density function of the  $\chi^2_6$  distribution (dotted line). Data are generated by the  $x_t$  model (left chart) and  $y_t$  model (right chart), while the  $y_t$  model is estimated.



data generating process it can be seen that the bootstrap distribution of the  $LR_{bh}$  test is approximately equal to the limiting  $\chi^2$  distribution. However, the skewness is somewhat larger than for the bootstrap (LR) and  $\chi^2$  distributions in Figure 4.4. Hence, inference based on the limiting distribution of the  $LR_{bh}$  statistic seems to be mildly biased in favor of the alternative relative to inference based on the bootstrap distribution. On the other hand, if data satisfies the term structure but *not* the cointegration constraints, the bootstrap distribution of the  $LR_{bh}$  test has shifted a significant distance to the right of the  $\chi^2$ . It should perhaps be emphasized that the  $\chi^2$  distribution is still valid under the null hypothesis and that this hypothesis is, by construction, generally only nearly consistent with the  $x_t$ -model. Hence, the  $LR_{bh}$  test seems fairly well equipped for rejecting the joint hypothesis of cointegration and the term structure model when the former is almost but not entirely true. For example, the p-values of the  $LR_{bh}$  test on Swedish interest rates in Table 4.5 indicate that the  $y_t$ -model is probably misspecified while the  $x_t$ -model is consistent with the observed value of this test.

Table 4.5: Bootstrap and  $\chi^2$  based p-values of tests of the term structure model when data is generated by a VAR(2) model.

estimated model		data model		$\chi^2$	(df)
		$x_t$	$y_t$		
$x_t$	LR	.013	.010	.008	(4)
	W	.013	.010	.004	(4)
$y_t$	LR	.043	.039	.040	(4)
	W	.043	.046	.029	(4)
$y_t$	$LR_{bh}$	.257	.024	.014	(6)

**Note:** The p-values for the bootstrap distributions are based on the percentile method (cf. Efron [23, Chapter 10]).

Let me summarize the discussion by referring to the three questions stated at the beginning of this subsection. First, since the  $y_t$ -model is conditioned on a set of linear restrictions (suggested but not implied by theory) it is reasonable to propose that these constraints should be tested along with those implied by the term structure model. If the  $LR_{bh}$  test rejects the joint hypothesis of cointegration and the term structure model,

the  $x_t$ -model should be employed when testing the latter restrictions in isolation. On the other hand, if the joint hypothesis is not rejected, the results indicate that it may not matter which data generating model is used for testing the theory. Second, since the bootstrap distributions of the LR tests are slightly less skewed to the right of the  $\chi^2$  than the bootstrap distributions of the Wald tests, it follows that the choice of test statistic may be important. Third, the limiting  $\chi^2$  distribution is very similar to most bootstrap distributions. Furthermore, when the  $x_t$ -model generates data and the  $y_t$ -model is used for estimation, i.e., generally a model misspecification, the bootstrap distributions of the test statistics have not shifted away from the  $\chi^2$  distribution by any notable distance as long as *only* the term structure restrictions are investigated. Hence,  $\chi^2$  based inference seems fairly robust with respect to such model misspecification.

## 4.5 Concluding Remarks

In this chapter I have analysed and tested a rational expectations model of the term structure of interest rates. Based on the local expectations hypothesis and a linear approximation of the one-period ahead holding period yield, the term structure model implies that the long term (consol) yield can be expressed as a weighted average of the current and expected future short term yields plus a constant term premium. The weights on the short term rates sum to one and are exponentially declining in the forecast horizon. Equivalently, the term structure model specifies that the spread between the long term and short term yield is a weighted average of expected future changes in the short term rate plus a constant term premium. When the economic hypothesis is connected with a bivariate finite order vector autoregression we find that either a levels or a cointegrated data generating model can be used. Furthermore, while the latter model is only valid under the form of cointegration suggested by the term structure hypothesis, the former model is also consistent with, e.g., jointly wide sense stationary interest rates. More importantly, if interest rates are modelled as being cointegrated of order (1,1) with a stationary spread, standard inference can be carried out within either data generating model. Also, if it is assumed that interest rates are nonexplosive and the term structure model holds true, I have shown that the long term yield Granger-causes the short term yield. Should such a condition not be satisfied by unconstrained estimates of either vector autoregression, the term structure restrictions imply that interest rates are explosive, i.e., standard inference from tests of these restrictions is no longer applicable when we condition on the term structure model.

From the empirical analyses it is found that the choice of lag order seems to be especially important. For example, if the log or iterated log criteria determines the lag order, the term structure restrictions are consistent with Swedish data, while a lag order selection based on the Akaike criterion provides us with the opposite conclusion. To deal with this I suggest that information criteria may be used to choose an upper and a lower bound for the lag order and that lag order tests help us in selecting a suitable lag order within this set.

To examine whether data are cointegrated, I propose some simple LR tests. In the present setting these tests have a limiting  $\chi^2$  distribution. Moreover, one of the statistics, denoted by  $LR_{bh}$ , allows us to test the cointegration and the term structure restrictions jointly. In connection with inference from tests of the term structure model on Swedish

data, we find that it can generally be rejected. However, it is not irrelevant whether the levels or the cointegrated vector autoregression represents the data generating process. In general, the cointegrated vector autoregression is not strongly at odds with the term structure model, whereas the levels model is. In view of the fact that a discount factor  $\rho$  being approximately equal to one and the term structure restrictions imply that jointly wide sense stationary interest rates are *nearly* cointegrated, it is perhaps not surprising that the  $y_t$ -model is somewhat more in line with the term structure model than the  $x_t$ -model is. Furthermore, the  $LR_{bh}$  test suggests that either cointegration or the term structure restrictions are not supported by data.

Finally, I have used bootstrap resampling to obtain alternative estimates (to the limiting  $\chi^2$ ) of the unknown small sample distributions of the LR and Wald tests. It is then found that if we only test the term structure restrictions, inference does not seem to be greatly effected by which model actually generates data. In particular, the tests seem to be quite robust with respect to misspecification of the cointegration restrictions. However, once the joint hypothesis of cointegration and the term structure model is tested, it can be concluded that if the  $LR_{bh}$  test rejects this hypothesis, the levels model should be employed when testing the term structure model. Furthermore, the  $\chi^2$  distribution performs quite well in relation to the bootstrap distributions as an estimate of the small sample uncertainty. Moreover, the  $\chi^2$  and the bootstrap LR distributions are less different (in terms of p-values) than the  $\chi^2$  and the bootstrap Wald distributions, thereby suggesting that the choice of test statistic may be important when the  $\chi^2$  distribution is considered for inference from statistical analysis of the term structure model.

Natural extensions of this study are currently being pursued by Ossian Ekdahl and myself. First, the i.i.d. assumption in the bootstrap analysis may be dropped and replaced with the assumption that  $\varepsilon_t$  is conditionally heteroskedastic and model conditional heteroskedasticity as a multivariate GARCH process (cf. Bollerslev [11]). This makes it possible to consider, e.g., the  $W^*$  statistic in a bootstrap study.

Second, time varying term premia are also of interest. Engle, Lilien and Robins [27] study (single equation) term structure tests of, e.g., a relationship similar to that examined by Mankiw and Summers [66] and assume that the residual is generated by an ARCH-M process. This time series model is based on the assumption that the conditional mean of  $\varepsilon_t$  is a function of its conditional variance. Accordingly, the model allows the term premium to be a function of volatility. In terms of a bivariate vector autoregression we aim to model  $\varepsilon_t$  as a multivariate GARCH-M process. To accomplish these two related extensions some new econometric tools must first be developed.

## Appendix: A Transformed Model for Cointegrated Interest Rates

The purpose of this appendix is to derive a transformed model of equation (4.9) which can be written in the form of equations (2.2) and (2.3) in Sims, Stock and Watson [92]. The transformed model (or transformed regression equation if we stick to the terminology in Sims, Stock and Watson) can then be used to show that all coefficients in (4.9) correspond to coefficients on mean zero stationary canonical regressors in the transformed model. In the terminology of Sims, Stock and Watson, a “canonical” regressor is a linear

transformation of the original regressors  $x_t$ . Furthermore, I shall show that lag order, Granger non-causality, cointegration, and term structure restrictions on the coefficients of (4.9) imply and are implied by a corresponding set of restrictions on the coefficients of the transformed model for interest rates which are cointegrated of order (1,1) with a stationary spread (see also, case 3 in Sims, Stock and Watson and the analysis in West [101]).

To derive the transformed model for  $x_t$ , let us begin by rewriting the vector autoregression in (4.9) such that

$$x_t = \mu + \sum_{j=1}^{p-1} A_j^* \Delta x_{t-j} + \left( \sum_{k=1}^p A_k \right) x_{t-1} + \varepsilon_t, \quad (4.27)$$

where

$$A_j^* := - \sum_{k=j+1}^p A_k \quad \forall j \in \{1, \dots, p-1\},$$

and  $\mu = \gamma\mu_0 = \gamma\alpha'\tilde{\mu} = A(1)\tilde{\mu}$ . The constraints on  $\mu$  ensure that  $x_t$  is not driven by any linear deterministic trends. It should be emphasized, however, that similar results as those below hold when  $\mu$  is not constrained as in the present case. Since the matrix  $A(1) = I_2 - \sum_{k=1}^p A_k$ , we find that (4.27) can be written as

$$x_t - \tilde{\mu} = \sum_{j=1}^{p-1} A_j^* \Delta x_{t-j} + (I_2 - A(1))(x_{t-1} - \tilde{\mu}) + \varepsilon_t. \quad (4.28)$$

If  $\{x_t\}_{t=1}^\infty$  is cointegrated of order (1,1) with 0, 1, or 2 cointegration vectors, then all elements of  $A_j^*$  for all  $j \in \{1, \dots, p-1\}$  are coefficients on mean zero stationary canonical regressors. This follows from the fact that  $\{\Delta x_t\}_{t=1}^\infty$  is jointly wide sense stationary and  $E[\Delta x_t] = 0$  for all  $t \geq 1$ . Hence, to show that all coefficients in (4.9) correspond to coefficients on mean zero stationary canonical regressors, it remains to consider the elements of  $(I_2 - A(1))$ .

To accomplish this, let us follow the notation in Sims, Stock and Watson and introduce the transformation matrix  $W$  which produces the canonical regressors  $\zeta_t := W(x_t - \tilde{\mu})$  and  $\zeta_t^i$  is the  $i$ :th element of  $\zeta_t$ . If there is one cointegration vector,  $W$  should be specified so that  $\zeta_t$  has one stationary canonical regressor and one nonstationary. Specifically, when the spread between the long and the short term yield is stationary, let

$$Wx_t = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} r_t \\ R_t \end{bmatrix} = \begin{bmatrix} r_t + R_t \\ R_t - r_t \end{bmatrix}. \quad (4.29)$$

The first element of  $\zeta_t$ , i.e.,  $\zeta_t^1$ , is thus dominated by a stochastic trend while the second element,  $\zeta_t^2$ , is stationary. This follows from the observation that the first row of  $W$  is orthogonal to the second row, and the second row is a cointegration vector. Furthermore, from, e.g., Engle and Granger [26] and Theorem 2.1, we know that  $\Delta x_t = C(L)\varepsilon_t$ , and

$$x_t - \tilde{\mu} = C(1)\xi_t + \tilde{C}(L)\varepsilon_t, \quad (4.30)$$

where  $\xi_t = \xi_{t-1} + \varepsilon_t$ ,  $\text{rank}[C(1)] = 1$ ,  $\alpha'C(1) = 0$ , and  $\{u_t\}_{t=1}^\infty$  (where  $u_t := \tilde{C}(L)\varepsilon_t$ ) is jointly wide sense stationary. Since  $\alpha' = [-1 \ 1]$  we know that  $(R_t - r_t)$  is not influenced

by the stochastic trend  $\xi_t$ . It can also be seen that  $\tilde{\mu} = [\bar{r} \ \bar{R}]'$  and that  $\mu_0 = \alpha' \tilde{\mu} = \bar{R} - \bar{r}$ . Accordingly, if the term structure model is consistent with data, then  $\mu_0 = \psi$ , the term premium.

Making use of equation (4.29), the transformed model for  $x_t$  can be written as

$$x_t - \tilde{\mu} = \sum_{j=1}^{p-1} A_j^* \Delta x_{t-j} + (I_2 - A(1)) W^{-1} e_2 \zeta_{t-1}^2 + (I_2 - A(1)) W^{-1} e_1 \zeta_{t-1}^1 + \varepsilon_t, \quad (4.31)$$

where  $e_i$  denotes the  $i$ :th column of  $I_2$ . From this equation it can be seen that since  $W^{-1} e_2 = [-.5 \ .5]'$  does *not* have a zero element, all elements of  $(I_2 - A(1))$  correspond to coefficients on the mean zero stationary canonical regressor  $\zeta_t^2$ . It may be noted that these elements also correspond to coefficients on the canonical regressor  $\zeta_t^1$  which is dominated by a stochastic trend. However, the rate of convergence of (least squares) estimates of the elements in  $(I_2 - A(1))$  is  $T^{1/2}$  as this rate is determined by the regressor which is associated with the slowest rate of convergence, i.e., the canonical regressor  $\zeta_t^2$ . For example, if interest rates are nonstationary in levels but not in first differences and there is no cointegration vector, the elements of  $(I_2 - A(1))$  would converge to some limiting distribution (see Sims, Stock and Watson [92, Theorem 1]) at a rate  $T$ , i.e., faster than when there exists a cointegration vector for the interest rates. Before I turn to the restrictions, note that the Sims, Stock and Watson "moving average" model (equations (2.2) and (2.3)) can be derived directly from (4.30) and (4.31). In particular, letting  $Z_t^1 := [\Delta x_t' \cdots \Delta x_{t-p+2}' \ \zeta_t^2]'$  and  $Z_t^3 := \zeta_t^1$  (the Sims, Stock and Watson notation is used), equation (2.2) in Sims, Stock and Watson is given by

$$Z_t = \begin{bmatrix} Z_t^1 \\ Z_t^3 \end{bmatrix},$$

whereas equation (2.3) may be written as

$$Z_t = \begin{bmatrix} F_{11}(L) & 0 \\ w_1 \tilde{C}(L) & w_1 C(1) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \xi_t \end{bmatrix}.$$

The vector  $w_1$  is the first row of  $W$ . Since  $w_1 \alpha' = 0$  and  $\alpha' C(1) = 0$ , we know that  $w_1 C(1) \neq 0$ , i.e.,  $Z_t^3$  is dominated by a stochastic trend.

The four main hypotheses being tested on the coefficients of equation (4.9) in section 4.4.2 are related to lag order, Granger-noncausality, cointegration, and term structure restrictions. The lag order tests are particularly simple to study. Suppose the null hypothesis is  $s$  versus  $p$  lags. In that case

$$H_0 : A_{s+1} = \cdots = A_p = 0.$$

Accordingly,

$$H_0^* : A_s^* = \cdots = A_{p-1}^* = 0,$$

is an equivalent set of restrictions. A Wald or LR test of the latter restrictions have a limiting  $\chi^2$  distribution, and since the hypotheses  $H_0$  and  $H_0^*$  are equivalent, a Wald or LR test of the former restrictions also has a limiting  $\chi^2$  distribution. Second, the Granger-noncausality test within the model (4.9) is based on the null hypothesis

$$H_0 : a_{ijk} = 0 \ \forall k \in \{1, \dots, p\}, \ i \neq j.$$

In terms of equation (4.31), an equivalent set of restrictions is

$$H_0^* : a_{ij}^* = 0 \text{ and } a(1)_{ij} = 0 \forall k \in \{1, \dots, p-1\}, i \neq j,$$

where  $a(1)_{ij}$  is the  $(i, j)$ :th element of  $A(1)$ . Third, the term structure model implies the hypothesis

$$H_0 : a_{211} = \frac{\rho - 1}{\rho}, a_{221} = \frac{1}{\rho}, \text{ and } a_{2jk} = 0 \forall k \in \{2, \dots, p\} \text{ and } j \in \{1, 2\},$$

when a test is conducted in equation (4.9). Equivalently, the term structure restrictions on the coefficients of the transformed model are given by

$$H_0^* : a_{2jk}^* = 0 \forall k \in \{1, \dots, p-1\} \text{ and } j \in \{1, 2\} \text{ and } a(1)_{21} = -a(1)_{22} = \frac{\rho - 1}{\rho}.$$

Fourth, if the cointegration and the term structure restrictions are examined jointly, in addition to the term structure constraints above we have that cointegration implies the following for the untransformed model:

$$H_0 : \sum_{k=1}^p (a_{11k} + a_{12k}) = 1.$$

The corresponding restrictions on the coefficients of (4.31) are those above for the term structure and

$$H_0^* : a(1)_{11} = -a(1)_{12},$$

for cointegration. Hence, all the restrictions being tested in the untransformed model (4.9) correspond to restrictions on the coefficients of mean zero stationary canonical regressors in the transformed model (4.31). Accordingly, the Wald and LR tests have their usual limiting  $\chi^2$  distribution even when interest rates are cointegrated of order (1,1) with a stationary spread. Finally, it may be noted that results similar to those above hold for all bivariate vector autoregressions where both elements of the cointegration vector are nonzero, i.e., for all cointegrated systems with one cointegration vector where none of the original regressors are jointly wide sense stationary (for more details, see West [101]).

# Bibliography

- [1] Ahmed, Shaghil, "Government Spending, the Balance of Trade, and the Terms of Trade in British History," *Journal of Monetary Economics* 20 (1987), 195–220.
- [2] Akaike, Hirotugu, "Information Theory and an Extension of the Maximum Likelihood Principle," In B.N. Petrov and F. Csáki, editors, *Second International Symposium on Information Theory*, pages 267–281, Akadémiai Kiado, Budapest, 1973.
- [3] Amemiya, Takeshi, *Advanced Econometrics*, Harvard University Press, Cambridge, 1985.
- [4] Baillie, Richard T., "Econometric Tests of Rationality and Market Efficiency," *Econometric Reviews*, 1989, Forthcoming.
- [5] Baillie, Richard T., "Inference in Dynamic Models containing 'Surprise' Variables," *Journal of Econometrics* 35 (1987), 101–117.
- [6] Baillie, Richard T., "Prediction from the Dynamic Simultaneous Equation Model with Vector Autoregressive Errors," *Econometrica* 49 (1981), 1331–1337.
- [7] Bergman, Michael and Lars Jonung, "The Business Cycle has not been Dampened? The Case of Sweden and the United States 1873–1988," Manuscript, University of Lund, 1989.
- [8] Bernanke, Ben S., "Alternative Explanations of the Money–Income Correlation," *Carnegie–Rochester Conference Series on Public Policy* 25 (1986), 49–100.
- [9] Berndt, Ernst R. and N. Eugene Savin, "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica* 45 (1977), 1263–1277.
- [10] Blanchard, Olivier J. and Danny Quah, "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review* 79 (1989), 655–673.
- [11] Bollerslev, Tim, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31 (1986), 307–327.
- [12] Campbell, John Y., "A Defense of Traditional Hypotheses about the Term Structure of Interest Rates," *Journal of Finance* 41 (1986), 183–193.
- [13] Campbell, John Y., "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," *Econometrica* 55 (1987), 1249–1273.

- [14] Campbell, John Y. and Robert J. Shiller, "Cointegration and Tests of Present Value Models," *Journal of Political Economy* 95 (1987), 1062–1088.
- [15] Campbell, John Y. and Robert J. Shiller, "Interpreting Cointegrated Models," *Journal of Economic Dynamics and Control* 12 (1988), 505–522.
- [16] Christiano, Lawrence J. and Martin Eichenbaum, "Unit Roots in Real G.N.P.: Do we know, and do we care?" NBER Working Paper No. 3130, 1989.
- [17] Cooley, Thomas F. and Stephen F. LeRoy, "Atheoretical Macroeconometrics: A Critique," *Journal of Monetary Economics* 16 (1985), 283–308.
- [18] Cox, John C., Ingersoll, Jr., Jonathan E. and Stephen A. Ross, "A Re-Examination of Traditional Hypotheses about the Term Structure of Interest Rates," *Journal of Finance* 36 (1981), 769–799.
- [19] Davidson, Russell and James G. MacKinnon, "Implicit Alternatives and the Local Power of Test Statistics," *Econometrica* 55 (1987), 1305–1329.
- [20] DeJong, David N. and Charles H. Whiteman, "Trends and Random Walks in Macroeconomic Time Series: A Reconsideration Based on the Likelihood Principle," Working Paper Series No. 89–04, University of Iowa, 1989.
- [21] Doan, Thomas A., Litterman, Robert B. and Christopher A. Sims, "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews* 3 (1984), 1–100, With comments by Thomas F. Cooley, John F. Geweke, Edmond Malinvaud, Angelo Melino, and Paul Newbold.
- [22] Dunsmuir, W. and Edward J. Hannan, "Vector Linear Time Series Models," *Advances in Applied Probability* 8 (1976), 339–364.
- [23] Efron, Bradley, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1982.
- [24] Ekdahl, Ossian, "Ex Ante Real Interest Rates," Research Paper No. 6422, Stockholm School of Economics, 1990.
- [25] Engle, Robert F., "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics*, pages 775–826, Elsevier Science Publishers, 1984.
- [26] Engle, Robert F. and Clive W.J. Granger, "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55 (1987), 251–276.
- [27] Engle, Robert F., Lilien, David M. and Russell P. Robins, "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica* 55 (1987), 391–407.
- [28] Engle, Robert F. and Byung Sam Yoo, "Forecasting and Testing in Co-Integrated Systems," *Journal of Econometrics* 35 (1987), 143–159.

- [29] Englund, Peter, Persson, Torsten and Lars E.O. Svensson, "Swedish Business Cycles, 1861–1988," Work in Progress, Institute for International Economic Studies, University of Stockholm, 1990.
- [30] Fama, Eugene F., "The Information in the Term Structure," *Journal of Financial Economics* 13 (1984), 509–528.
- [31] Fama, Eugene F., "Term Premiums in Bond Returns," *Journal of Financial Economics* 13 (1984), 529–546.
- [32] Froot, Kenneth A., "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Finance* 44 (1989), 283–305.
- [33] Geweke, John F. and Richard Meese, "Estimating Regression Models of Finite but Unknown Order," *International Economic Review* 22 (1981), 55–70.
- [34] Granger, Clive W.J. and Andrew A. Weiss, "Time Series Analysis of Error-Correction Models," In S. Karlin, editor, *Studies in Econometrics, Time Series, and Multivariate Statistics*, pages 255–278, Academic Press, New York, 1983.
- [35] Gregory, Allan W. and Michael R. Veall, "Formulating Wald Tests of Nonlinear Restrictions," *Econometrica* 53 (1985), 1465–1468.
- [36] Haldrup, Niels, "Tests for Unit Roots with a Maintained Trend when the True Data Generating Process is a Random Walk with Drift," Manuscript, Institute of Economics, University of Aarhus, Denmark, 1990.
- [37] Haldrup, Niels and Svend Hylleberg, "Unit Roots and Deterministic Trends, with yet another Comment on the Existence and Interpretation of a Unit Root in U.S. G.N.P.," Manuscript, Institute of Economics, University of Aarhus, Denmark, 1989.
- [38] Hannan, Edward J., *Multiple Time Series*, John Wiley, New York, 1970.
- [39] Hannan, Edward J. and B.G. Quinn, "The Determination of the Order of an Autoregression," *Journal of the Royal Statistical Society B* 41 (1979), 190–195.
- [40] Henderson, Harold V. and S.R. Searle, "Vec and Vech Operators for Matrices, with some uses in Jacobians and Multivariate Statistics," *Canadian Journal of Statistics* 7 (1979), 65–81.
- [41] Hjerpe, Riitta, *The Finnish Economy 1860–1985: Growth and Structural Change*, Bank of Finland, Helsinki, 1989.
- [42] Hylleberg, Svend and Grayham E. Mizon, "Cointegration and Error Correction Mechanisms," *Economic Journal* 99 (1989), 113–125.
- [43] Jacobson, Tor, "On the Determination of Lag Order in Vector Auto Regressions of Cointegrated Systems," Manuscript, Department of Statistics, University of Uppsala, 1990.
- [44] Johansen, Søren, "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 1990, Forthcoming.

- [45] Johansen, Søren, "Likelihood Based Inference on Cointegration: Theory and Applications," Lecture Notes, Institute of Mathematical Statistics, University of Copenhagen, 1989.
- [46] Johansen, Søren, "The Mathematical Structure of Error Correction Models," *Contemporary Mathematics* 80 (1988), 359–386.
- [47] Johansen, Søren, "The Power Function of the Likelihood Ratio Test for Cointegration," Preprint No. 8, Institute of Mathematical Statistics, University of Copenhagen, 1989.
- [48] Johansen, Søren, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12 (1988), 231–254.
- [49] Johansen, Søren and Katarina Juselius, "Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for U.K.," Preprint No. 1, Institute of Mathematical Statistics, University of Copenhagen, 1990.
- [50] Keating, John W., "Identifying VAR Models under Rational Expectations," *Journal of Monetary Economics* 25 (1990), 453–476.
- [51] King, Robert G., Plosser, Charles I. and Sergio T. Rebelo, "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics* 21 (1988), 195–232.
- [52] King, Robert G., Plosser, Charles I. and Sergio T. Rebelo, "Production, Growth and Business Cycles: II. New Directions," *Journal of Monetary Economics* 21 (1988), 309–341.
- [53] King, Robert G., Plosser, Charles I., Stock, James H. and Mark W. Watson, "Stochastic Trends and Economic Fluctuations," NBER Working Paper No. 2229, 1987.
- [54] Krantz, Olle and Carl-Axel Nilsson, *Swedish National Product 1861–1970: New Aspects on Methods and Measurement*, C.W.K. Gleerup/Liber Läromedel, Lund, 1975.
- [55] Litterman, Robert B., "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience," *Journal of Business and Economic Statistics* 4 (1986), 25–38.
- [56] Litzenberger, Robert H. and Jacques Rolfo, "An International Study of Tax Effects on Government Bonds," *Journal of Finance* 39 (1984), 1–22.
- [57] Lundvik, Petter, "Business Cycles in a Small Open Economy: Sweden 1871–1987," Manuscript, Institute for International Economic Studies, University of Stockholm, 1989.
- [58] Lütkepohl, Helmut, "Asymptotic Distribution of the Moving Average Coefficients of an Estimated Vector Autoregressive Process," *Econometric Theory* 4 (1988), 77–85.

- [59] Lütkepohl, Helmut, "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics* 72 (1990), 116–125.
- [60] Lütkepohl, Helmut, "A Note on the Asymptotic Distribution of Impulse Response Functions of estimated VAR Models with Orthogonal Residuals," *Journal of Econometrics* 42 (1989), 371–376.
- [61] Lütkepohl, Helmut and Donald S. Poskitt, "Estimating Orthogonal Impulse Responses via Vector Autoregressive Models," Manuscript, Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität, Kiel, West Germany, 1990.
- [62] Lütkepohl, Helmut and Hans-Eggert Reimers, "Impulse Response Analysis of Co-Integrated Systems with an Investigation of German Money Demand," Manuscript, Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität, Kiel, West Germany, 1990.
- [63] Magnus, Jan R. and Heinz Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley, Chichester, 1988.
- [64] Magnus, Jan R. and Heinz Neudecker, "Matrix Differential Calculus with Applications to Simple, Hadamard, and Kronecker Products," *Journal of Mathematical Psychology* 29 (1985), 474–492.
- [65] Magnus, Jan R. and Heinz Neudecker, "Symmetry, 0–1 Matrices and Jacobians: A Review," *Econometric Theory* 2 (1986), 157–190.
- [66] Mankiw, N. Gregory and Lawrence H. Summers, "Do Long-Term Interest Rates Overreact to Short-Term Interest Rates," *Brookings Papers on Economic Activity* 1 (1984), 223–242.
- [67] Neudecker, Heinz, "On Jacobians of Transformations with Skew-Symmetric, Strictly (Lower) Triangular or Diagonal Matrix Arguments," *Linear and Multilinear Algebra* 14 (1983), 271–295.
- [68] Nickelsburg, Gerald, "Small-Sample Properties of Dimensionality Statistics for Fitting VAR Models to Aggregate Economic Data," *Journal of Econometrics* 28 (1985), 183–192.
- [69] Oberhofer, W. and Jan Kmenta, "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models," *Econometrica* 42 (1974), 579–590.
- [70] Ohlsson, Lennart, *Utrikeshandeln och den ekonomiska tillväxten i Sverige 1871–1966* (The Foreign Trade and the Economic Growth in Sweden 1871–1966), Industriens Utredningsinstitut/Almqvist & Wiksell, Stockholm, 1969.
- [71] Paulsen, Jostein, "Order Determination of Multivariate Autoregressive Time Series with Unit Roots," *Journal of Time Series Analysis* 5 (1984), 115–127.

- [72] Perron, Pierre, "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach," *Journal of Economic Dynamics and Control* 12 (1988), 297–332.
- [73] Persson, Torsten and Lars E.O. Svensson, *New Methods in the Swedish Medium Term Survey*, Allmänna Förlaget, Stockholm, 1987, Supplement No. 3 to LU87.
- [74] Quinn, B.G., "Order Determination for a Multivariate Autoregression," *Journal of the Royal Statistical Society B* 42 (1980), 182–185.
- [75] Rao, C. Radhakrishna, *Linear Statistical Inference and Its Applications*, John Wiley, New York, 1973.
- [76] Reinsel, Gregory C., "Asymptotic Properties of Prediction Errors for the Multivariate Autoregressive Model Using Estimated Parameters," *Journal of the Royal Statistical Society B* 42 (1980), 328–333.
- [77] Reinsel, Gregory C. and Richard A. Lewis, "Prediction Mean Square Error for Non-Stationary Multivariate Time Series Using Estimated Parameters," *Economic Letters* 24 (1987), 57–61.
- [78] Rose, Andrew K., "Is the Real Interest Rate Stable?" *Journal of Finance* 44 (1989), 1095–1112.
- [79] Rothenberg, Thomas J., *Efficient Estimation with A Priori Information*, Yale University Press, New Haven, 1973.
- [80] Runkle, David E., "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics* 5 (1987), 437–454, With comments by Christopher A. Sims, Olivier J. Blanchard, and Mark W. Watson.
- [81] Sargent, Thomas J., "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics* 5 (1979), 133–143.
- [82] Schmidt, Peter, "The Asymptotic Distribution of Dynamic Multipliers," *Econometrica* 41 (1973), 161–164.
- [83] Schmidt, Peter, "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model," *Econometrica* 42 (1974), 303–309.
- [84] Schwarz, Gideon, "Estimating the Dimension of a Model," *Annals of Statistics* 6 (1978), 461–464.
- [85] Serfling, Robert J., *Approximation Theorems of Mathematical Statistics*, John Wiley, New York, 1980.
- [86] Shapiro, Matthew D. and Mark W. Watson, "Sources of Business Cycle Fluctuations," In Stanley Fisher, editor, *NBER Macroeconomics Annual*, pages 111–148, MIT Press, 1988.

- [87] Shea, Gary S., "Qualms About the Linearized Expectations Hypothesis and Variance-Bounds Studies of the Interest Rate Term Structure," Manuscript, Department of Finance, Pennsylvania State University, 1989.
- [88] Shibata, R., "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion," *Biometrika* 63 (1976), 117-126.
- [89] Shiller, Robert J., "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," *Journal of Political Economy* 87 (1979), 1190-1219.
- [90] Sims, Christopher A., "Macroeconomics and Reality," *Econometrica* 48 (1980), 1-48.
- [91] Sims, Christopher A., "Policy Analysis with Econometric Models," *Brookings Papers on Economic Activity* 1 (1982), 107-164.
- [92] Sims, Christopher A., Stock, James H. and Mark W. Watson, "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58 (1990), 113-144.
- [93] Sims, Christopher A. and Harald Uhlig, "Understanding Unit Rooters: A Helicopter Tour," Discussion Paper No. 4, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, 1988.
- [94] Stock, James H., "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica* 55 (1987), 1035-1056.
- [95] Stock, James H. and Mark W. Watson, "Testing for Common Trends," *Journal of the American Statistical Association* 83 (1988), 1097-1107.
- [96] Stock, James H. and Mark W. Watson, "Variable Trends in Economic Time Series," *Journal of Economic Perspectives* 2 (1988), 147-174.
- [97] Walsh, Carl E., "The Impact of Monetary Targeting in the United States: 1976-1984," NBER Working Paper No. 2384, 1987.
- [98] Warne, Anders, "Cointegration, Forecasting, and Some Linear Rational Expectations Models," Research Paper No. 6395, Stockholm School of Economics, 1989.
- [99] Warne, Anders, "Maximum Likelihood Estimation of a Vector Autoregression subject to Exact Nonlinear Cross Equation Restrictions," Paper in progress, Stockholm School of Economics, 1990.
- [100] Warne, Anders and Anders Vredin, "Current Account and Macroeconomic Fluctuations: The Case of Sweden," Manuscript, Stockholm School of Economics and Trade Union Institute for Economic Research (FIEF), 1990.
- [101] West, Kenneth D., "Asymptotic Normality, when Regressors have a Unit Root," *Econometrica* 56 (1988), 1397-1417.
- [102] White, Halbert, *Asymptotic Theory for Econometricians*, Academic Press, London, 1984.

- [103] White, Halbert, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50 (1982), 1–25.
- [104] Yamamoto, Taku, "Predictions of Multivariate Autoregressive–Moving Average Models," *Biometrika* 68 (1981), 485–492.

# EFI - publikationer sedan 1987

---

## 1987

---

Andersson, T., Ternström, B., Private Foreign Investments and Welfare Effects: A comparative study of five countries in southeast Asia. Research report

Andersson, T., Ternström, B., External Capital and Social Welfare in South-East Asia. Research report

Benndorf, H., Marknadsföringsplanering och samordning mellan företag i industriella system. EFI/MTC

Bergman, L., Mäler, K-G., Ståhl, I., Överlåtelsebara utsläppsrätter. En studie av kolväteutsläpp i Göteborg. Research report

Björkegren, D., Mot en kognitiv organisationsteori. Research report

Claesson, K., Effektiviteten på Stockholms Fondbörs

Davidsson, P. Growth Willingness in Small Firms. Entrepreneurship - and after? Research report

Engshagen, I., Finansiella nyckeltal för koncern versus koncernbolag. En studie om finansiella måttal. Research report

Fredriksson, O., Holmlöv, P.G., Julander, C-R., Distribution av varor och tjänster i informationssamhället.

Hagstedt, P., Sponsring - mer än marknadsföring. EFI/MTC

Järnhäll, B., On the Formulation and Estimation of Models of Open Economies. Research report

Kylén, B., Digitalkartans ekonomi. Samhällsekonomska modeller för strategiska val. Delrapport 2 i "Nordisk Kvantif"

Lundgren, S., Elpriser: Principer och praktik. Research report

Schwarz, B., Hederstierna, A., Socialbidragens utveckling - orsaker eller samvariationer? Research report

Schwartz, E., Begreppet federativ organisation belyst i ett organisationsteoretiskt, juridiskt och empiriskt perspektiv. Research report.

Westlund, A.H. Öhlén, S., Business Cycle Forecasting in Sweden; A problem analysis. Research report

---

## 1988

---

Andréasson, I-M., Costs of Controls on Farmers' Use of Nitrogen. A study applied to Gotland.

Björkegren, D., Från slutet till öppen kunskapsproduktion. Delrapport 3 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björkegren, D., Från tavistock till Human-relation. Delrapport 4 i forskningsprojektet Lärande och tänkande i organisationer. Research report

Björklund, L., Internationell projekt-försäljning. En studie av ad hoc-samverkan mellan företag vid internationella projekt. Research report

Bojö, J., Mäler, K-G., Unemo, L., Economic Analysis of Environmental Consequences of Development Projects. Research report

Brynell, K., Davidsson, P., Det nya småföretagandet? - en empirisk jämförelse mellan små high-tech företag och konventionella småföretag. Research report

Dahlgren, G., Witt, P., Ledning av fusionsförlopp. En analys av bildandet av Ericsson Information Systems AB.

Forsell, A., Från traditionell till modern sparbank. Idé organisation och verksamhet i omvandling. Research report

Hirdman, V., The Stability and the Interest Sensitivity of Swedish Short Term Capital Flows. Research report

Hultén, S., Vad bestämmer de svenska exportmarknadsandelarnas utveckling?

Häckner, J., Biobränslenas konkurrenskraft i ett framtida perspektiv. Research report

Jennergren, P., Näslund, B., If the Supreme Court had Known Option Theory: A Reconsideration of the Gimo Case. Research report

Jennergren, P., Näslund, B., The Gimo Corporation Revisited: An Exercise in Finance Theory. Research report

Jennergren, P., Näslund, B., Valuation of Debt and Equity Through One- and Two-State Contingent Claims Models - with an Application to a Swedish Court Case. Research report

Jonung, L., Laidler, D., Are Perceptions of Inflation Rational? Some Evidence for Sweden. Research report

Lagerstam, C., Business Intelligence. Teori samt empirisk studie av elektronik- och bioteknikbranschen i Japan. Research report

Liljegren, G., Interdependens och dynamik i långsiktiga kundrelationer. Industriell försäljning i nätverksperspektiv.

Näslund, B., Några synpunkter på börsfallet i oktober 1987. Research report

Olsson, C., The Cost-Effectiveness of Different Strategies Aimed at Reducing the Amount of Sulphur Deposition in Europe. Research report

Philips, Å., Eldsjälar. En studie av aktörsskap i arbetsorganisatoriskt utvecklingsarbete.

Sellstedt, B., Produktionsstrategier. En diskussion med utgångspunkt från litteraturen. Research report

Skogsvik, K., Prognos av finansiell kris med redovisningsmått. En jämförelse mellan traditionell och inflationsjusterad redovisning.

Wahlund, R., Skatteomläggningen 1983 - 1985. En studie av några konsekvenser. Research report

Wahlund, R., Varför och hur olika svenska hushåll sparar. Research report

Vredin, A., Macroeconomic Policies and the Balance of Payments.

Åkerman, J., Economic Valuation of Risk Reduction: The Case of In-Door Radon. Research report.

---

## 1989

---

Andersson, T., Foreign Direct Investment in Competing Host Countries. A Study of Taxation and Nationalization.

Björkegren, D., Skönhetens uppfinnare. Research report

Björkegren, D., Hur organisationer lär. Studentlitteratur.

Blomström, M., Transnational Corporations and Manufacturing Exports from Developing Countries. Research report.

Carlsson, A., Estimates of the Costs of Emission control in the Swedish Energy Sector. Research report

Blomström, M., Foreign Investment and Spillovers. London; Routledge

Bordo, M.D., Jonung, L., The Long-Run Behavior of Velocity: The Institutional Approach Revisited. Research report

Davidsson, P., Continued Entrepreneurship and Small Firm Growth.

DeJuan, A., Fiscal Attitudes and Behavior. A study of 16 - 35 years old Swedish citizens. Research report

Edlund, P.-O., Preliminary Estimation of Transfer Function Weights. A Two-Step Regression Approach.

Fridman, B., Östman, L., eds. Accounting Development - some perspectives. In honour of Sven-Erik Johansson.

Gadde, L.-E., Håkansson, H., Öberg, M., Stability and Change in Automobile Distribution. Research report

Glader, M., Datorer i småföretag. Tel-dok

Jakobsson, B., Konsten att reagera. Intressen, institutioner och näringspolitik. Carlssons Bokförlag

Jonung, L., The Economics of Private Money. Private Bank Notes in Sweden 1831 - 1902. Research report

Jonung, L., Batchelor, R.A., Confidence about Inflation Forecasts: Tests of Variance Rationality. Research report.

Kylén, B., Hur företagschefer beslutar innan de blir överraskade. Ett försök till förklaring av svarsmönster i svagsignalsituationer.

Lagerstam, C., Jämförelse av skydd med samlad valuta/aktieoption och skydd med separata aktie- och valutoptioner samt härledning av optionspriser.

Research report

Lagerstam, C., On the Pricing and Valuation of Forwards and Options on Futures and their Relationship. Research report.

Larsson, B., Koncernföretaget. Ägarorganisationer eller organisation för ägare?

Lindberg, C., Holmlöv, P.G., Wärneryd K-E., Telefaxen och användarna. Teldok

Löwstedt, J., Föreställningar, ny teknik och förändring. Tre organisationsprocesser ur ett kognitivt aktörsperspektiv. Doxa Förlag

Löwstedt, J., (red) Organisation och teknikförändring. Studentlitteratur.

Tollgerdt-Andersson, I., Ledarskapsteori, företagsklimat och bedömningsmetoder.

Schuster, W., Ägandeformens betydelse för ett företag - en studie av ICA-rörelsen.

Schwartz, B., Företaget som medborgare. Samhällskontakter och reklam som legitimeringsinstrument. Research report

Spets, G., Vägen till nej. Anade och oanade konsekvenser av en OS-satsning. Research report.

Stymne, B., Information Technology and Competence Formation in the Swedish Service Sector. IMIT/EFI

Brunsson, N., Forsell, A., Winberg, H., Reform som tradition. Administrativa reformer i Statens Järnvägar.

Drottz-Sjöberg, B-M., Interests in Humanities, Social Science and Natural Science. A Study of High School Students.

Eklöf, J.A., Macro Economic Planning with Quantitative Techniques - and Effects of Varying Data Quality. Research Report.

Hesselman, A., Ett lokalt ekonomisystem. En fallstudie inom byggbranschen. Research Report.

Jennergren, L.P., Näslund, B., Models for the Valuation of International Convertible Bonds. Research Report.

Jonung, L., Introduction and Summary to The Stockholm School of Economics Revisited. Research Report.

Lagerstam, C., Hedging of Contracts, Anticipated Positions, and Tender Offers.

Patrickson, A., Essays in the Latin American Fertilizer Industry.

Steiner, L., Ledningsfunktionen i tillväxtföretag. Ledningsteamens sammansättning och funktion i tillverkande företag inom informationsteknologiindustrin.

Wärneryd, K., Economic Conventions. Essays in Institutional Evolution.

---

## 1990

---

Bojö, J., Economic Analysis of Agricultural Development Projects. A Case Study from Lesotho. Research Report.





